



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(11) PI 0903501-0 B1



(22) Data do Depósito: 23/03/2009

(45) Data de Concessão: 24/09/2020

(54) Título: CORREÇÃO DE PITCH DE ALTA QUALIDADE USANDO PROCESSAMENTO DE ÁUDIO

(51) Int.Cl.: G10L 19/02.

(52) CPC: G10L 19/0212; G10L 19/022.

(30) Prioridade Unionista: 04/04/2008 US 61/042,314; 08/12/2008 EP 08021298.8.

(73) Titular(es): FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V..

(72) Inventor(es): BAYER, STEFAN; DISCH, SASCHA; EDLER, BERND; FUCHS, GUILLAUME; GEIGER, RALF; KRAEMER, ULRICH; MULTRUS, MARKUS; NEUENDORF, MAX; POPP, HARALD; SCHULLER, GERALD.

(86) Pedido PCT: PCT EP2009002118 de 23/03/2009

(87) Publicação PCT: WO 2009/121499 de 08/10/2009

(85) Data do Início da Fase Nacional: 28/12/2009

(57) Resumo: CORREÇÃO DE PITCH DE ALTA QUALIDADE USANDO PROCESSAMENTO DE ÁUDIO É gerada uma representação processada de um sinal de áudio tendo uma sequência de frames pela amostragem do sinal de áudio dentro de um primeiro e um segundo frame da sequência de frames, c) segundo frame seguindo o primeiro frame, a amostragem usando informações de um contorno de pitch do primeiro e do segundo frame para obter uma primeira representação amostrada. O sinal de áudio é amostrado dentro do segundo e do terceiro frame, o terceiro frame seguindo o segundo frame na sequência de frames. A amostragem usa as informações no contorno de pitch do segundo frame e informações em um contorno de pitch do terceiro frame para obter uma segunda representação amostrada. A primeira janela de escalagem é obtida da primeira representação amostrada e uma segunda janela de escalagem é obtida da segunda representação amostrada, as janelas de escalagem dependendo das amostragens aplicadas para obter as primeiras representações amostradas ou a segunda representação amostrada.

"CORREÇÃO DE PITCH DE ALTA QUALIDADE USANDO PROCESSAMENTO DE ÁUDIO"

Campo da Invenção

Várias configurações da presente invenção se referem a processadores de áudio para a geração de uma representação processada de um sinal de áudio em frames usando amostragem dependente de pitch e da reamostragem dos sinais.

Histórico da Invenção e Técnica Anterior

São geralmente usadas transformadas sobrepostas moduladas em seno ou co-seno correspondendo aos bancos de filtros modulados em aplicações de codificação de fonte devido às suas propriedades de compactação de energia. Isto é, para tons harmônicos com frequências fundamentais constantes (pitch), concentram a energia do sinal em um baixo número de componentes espectrais (sub-bandas), o que leva a representações de sinais eficientes. Geralmente, o pitch de um sinal será entendido como sendo a menor frequência dominante distinguível do espectro do sinal. No modelo de fala comum, o pitch é a frequência do sinal de excitação modulado pela garganta humana. Se somente existir uma única frequência fundamental, o espectro seria extremamente simples, compreendendo somente a frequência fundamental e os sobretons. Esse espectro poderia ser codificado de forma altamente eficiente. Para sinais com variação de pitch, entretanto, a energia correspondente a cada componente harmônico é distribuído em vários coeficientes de transformadas, levando assim a uma redução da eficiência de codificação.

Poder-se-ia tentar melhorar a eficiência de codificação dos sinais com variação de pitch, primeiro criando um

sinal tempo discreto com um pitch virtualmente constante. Para que
 isso seja obtido, a taxa de amostragem poderia variar
 proporcionalmente ao pitch. Isto é, poder-se-ia reamostrar todo o
 sinal antes da aplicação da transformada, de maneira que o pitch
 5 seja o mais constante possível dentro de toda a duração do sinal.
 Isto pode ser feito por amostragem não equidistante, onde os
 intervalos de amostragem se adaptam localmente e são escolhidos de
 maneira que o sinal reamostrado, quando interpretado em termos de
 amostras equidistantes, tenha um contorno de pitch mais próximo de
 10 um pitch médio comum que o sinal original. Neste sentido, o
 contorno de pitch será entendido como sendo a variação local do
 pitch. A variação local poderia, por exemplo, ser parametrizada
 como uma função do tempo ou do número de amostras.

De forma equivalente, esta operação poderia ser
 15 vista como um *reescalagem* do eixo do tempo de um sinal amostrado
 ou de um sinal contínuo antes de uma amostragem equidistante. Essa
 transformada de tempo é também conhecida como *warping*. Aplicando
 uma transformada de frequência a um sinal que tenha sido pré-
 processado para chegar a um pitch quase constante, poderia
 20 aproximar a eficiência de codificação da eficiência que pode ser
 obtida para um sinal tendo um pitch genericamente constante.

A abordagem anterior, entretanto, tem várias
 desvantagens. Primeiro, uma variação da taxa de amostragem em uma
 ampla faixa, como necessário para o processamento do sinal
 25 completo, poderia levar a uma largura de banda do sinal muito
 variável devido ao teorema de amostragem. Segundo, cada bloco de
 coeficientes de transformadas representando um número fixo de
 amostras de entrada representaria então um segmento de tempo com

duração variável no sinal original. Isso tornaria as aplicações com retardo limitado de codificação praticamente impossíveis e, além disso, provocaria dificuldades de sincronização.

É proposto outro método pelos requerentes do pedido internacional de patente 2007/051548. Os autores propõem um método para fazer o *warping* em base de *frames*. Entretanto, isto é obtido pela introdução de restrições indesejáveis nos contornos aplicáveis de *warp*.

Portanto, existe a necessidade de outras abordagens para aumentar a eficiência de codificação, mantendo ao mesmo tempo a alta qualidade dos sinais codificados e decodificados de áudio.

Sumário da Invenção

Várias configurações da presente invenção permitem um aumento na eficiência de codificação por meio da realização de uma transformação local do sinal dentro de cada bloco de sinais (áudio frame) para o provimento de um pitch (virtualmente) constante na duração de cada bloco de entrada, contribuindo para um conjunto de coeficientes de transformadas em uma transformada em bloco. Esse bloco de entrada pode, por exemplo ser criado por dois *frames* consecutivos de um sinal de áudio quando é usada uma transformada co-seno discreta modificada como transformação no domínio da frequência.

Ao usar uma transformada sobreposta modulada, como a transformada co-seno discreta modificada (MDCT), dois blocos sucessivos entram na sobreposição da transformada no domínio da frequência para permitir um *cross-fade* do sinal nos limites do bloco, de forma a suprimir problemas audíveis do

processamento em bloco. É evitado um aumento do número de coeficientes de transformadas comparado com uma transformada não sobreposta por meio da amostragem crítica. Em MDCT, a aplicação da transformada para frente e para trás a um bloco de entrada, entretanto, não conduz à sua total reconstrução já que, devido à amostragem crítica, são introduzidos problemas no sinal reconstruído. A diferença entre o bloco de entrada e o sinal da transformada para frente e para trás é normalmente denominado de "aliasing no domínio do tempo" [aliasing no domínio do tempo].

Pela sobreposição dos blocos reconstruídos de uma metade da largura do bloco após a reconstrução e somando as amostras sobrepostas, o sinal de entrada pode, ainda assim, ser perfeitamente reconstruído no esquema MDCT. De acordo com algumas configurações, esta propriedade da transformada co-seno direta modificada pode ser mantida mesmo quando o sinal subjacente tiver *warp* de tempo em base de blocos (que é equivalente à aplicação de taxas de amostragem de adaptação local).

Como descrito anteriormente, a amostragem com taxas de amostragem de adaptação local (uma taxa de amostragem variável) pode ser vista como amostragem uniforme em uma escala de tempo em *warp*. Com esta visão, uma compactação da escala de tempo antes da amostragem leva a uma taxa de amostragem menos efetiva, aumentando o esticamento da taxa efetiva de amostragem do sinal subjacente.

Considerando uma transformada de frequência ou outra transformada, que use sobreposição e some na reconstrução para compensar possíveis problemas, o cancelamento do *aliasing* no domínio do tempo ainda funciona se o mesmo *warping* (correção de

pitch) for aplicado na região de sobreposição de dois blocos sucessivos. Assim, o sinal original pode ser reconstruído após a inversão do *warping*. Isto também é verdade quando diferentes taxas locais de amostragem são escolhidas nos dois blocos de transformadas sobrepostas, já que o *aliasing* no domínio do tempo do correspondente sinal de tempo contínuo ainda se cancela, dado que o teorema de amostragem é obedecido.

Em algumas configurações, na taxa de amostragem após o *warping* de tempo o sinal dentro de cada bloco de transformada é selecionado individualmente para cada bloco. Isto tem o efeito de um número fixo de amostras ainda representar um segmento de duração fixa no sinal de entrada. Além disso, pode ser usado um *amostrador*, que faz a amostragem do sinal de áudio dentro dos blocos de transformadas sobrepostas usando as informações sobre o contorno de pitch do sinal, de maneira que a porção de sinal sobreposta de uma primeira representação amostrada e de uma segunda representação amostrada tem contorno de pitch similar ou idêntico em cada uma das representações amostradas. O contorno de pitch ou as informações sobre o contorno de pitch usadas na amostragem podem ser obtidas arbitrariamente, enquanto houver uma interrelação não ambígua entre as informações sobre o contorno de pitch (o contorno de pitch) e o pitch do sinal. As informações sobre o contorno de pitch usadas, por exemplo, podem ser do pitch absoluto, do pitch relativo (a alteração do pitch), uma fração do pitch absoluto ou uma função que depende de forma não ambígua do pitch. A escolha das informações sobre o contorno de pitch como acima indicado, a porção da primeira representação amostrada que corresponde ao segundo frame tem um contorno de pitch similar ao

contorno de pitch da porção da segunda representação amostrada que corresponde ao segundo frame. A similaridade pode, por exemplo, ser que os valores do pitch das correspondentes porções de sinal tenham uma razão mais ou menos constante, isto é, uma razão dentro de uma faixa pré-determinada de tolerância. A amostragem pode assim ser feita de maneira que a porção da primeira representação amostrada que corresponde ao segundo frame tenha um contorno de pitch dentro de uma faixa pré-determinada de tolerância de um contorno de pitch da porção da segunda representação amostrada que corresponde ao segundo frame.

Como o sinal dentro do bloco de transformadas pode ser reamostrado com diferentes frequências de amostragem ou intervalos de amostragem, são criados blocos de entrada que podem ser codificados de forma eficiente por um subsequente algoritmo de codificação de transformadas. Isto pode ser feito enquanto, ao mesmo tempo, são aplicadas as informações obtidas no contorno de pitch sem quaisquer outras restrições enquanto o contorno de pitch for contínuo.

Mesmo que não seja obtida nenhuma alteração relativa de pitch dentro de um único bloco de entrada, o contorno de pitch pode ser mantido constante dentro e nos limites daqueles intervalos de sinal ou blocos de sinais que não possuam alteração obtível de pitch. Isto pode ser vantajoso quando falha ou é errôneo o rastreamento de pitch, que pode ser o caso em sinais complexos. Mesmo neste caso, o ajuste de pitch ou a reamostragem anterior à codificação da transformada não proporciona nenhum problema adicional.

A amostragem independente dentro dos blocos de

entrada pode ser feita usando janelas de transformadas especiais
 (janelas de escalagem) aplicadas antes ou durante a transformada
 no domínio da frequência. De acordo com algumas configurações,
 essas janelas de escalagem dependem do contorno de pitch dos
 5 frames associados ao bloco de transformadas. Em termos gerais, as
 janelas de escalagem dependem da amostragem aplicada para a
 obtenção da primeira representação amostrada ou da segunda
 representação amostrada. Isto é, a janela de escalagem da primeira
 representação amostrada pode depender da amostragem aplicada para
 10 obter somente a primeira janela de escalagem, na amostragem
 aplicada para obter somente a segunda janela de escalagem ou
 ambas, a amostragem aplicada para obter a primeira janela de
 escalagem e a amostragem aplicada para obter a segunda janela de
 escalagem. O mesmo se aplica, com as devidas alterações, à janela
 15 de escalagem da segunda representação amostrada.

Isto dá a possibilidade de garantia de que não
 mais que dois blocos subsequentes se sobreponham a qualquer
 momento durante a sobreposição e somem a reconstrução, de maneira
 que o cancelamento do *aliasing* no domínio do tempo seja possível.

20 Em particular, as janelas de escalagem da
 transformada são, em certas configurações, criadas de maneira a
 que possam ter formas diferentes dentro de cada uma das duas
 metades de cada bloco de transformada. Isto é possível enquanto
 cada metade de janela obedeça a condição de cancelamento de
 25 *aliasing* em conjunto com a metade da janela do bloco vizinho
 dentro do intervalo comum de sobreposição.

Apesar de as taxas de amostragem dos dois blocos
 de sobreposição poderem ser diferentes (valores diferentes dos

sinais subjacentes de áudio correspondem a amostras idênticas), o mesmo número de amostras pode agora corresponder a diferentes porções do sinal (formas de sinal). Entretanto, a exigência anterior pode ser preenchida reduzindo o comprimento de transição (amostras) de um bloco com uma taxa de amostragem menos efetiva que seu bloco associado de sobreposição. Em outras palavras, pode ser usada uma calculadora de janela transformada ou um método para calcular as janelas de escalagem, o que proporciona janelas de escalagem com um número idêntico de amostras para cada bloco de entrada. Entretanto, o número de amostras usadas para fazer o *fade out* do primeiro bloco de entrada pode ser diferente do número de amostras usado para fazer o *fade* no segundo bloco de entrada. Assim, usando janelas de escalagem para as representações amostradas de blocos sobrepostos de entrada (a primeira representação amostrada e a segunda representação amostrada), que dependem da amostragem aplicada aos blocos de entrada, permite uma diferente amostragem dentro dos blocos sobrepostos de entrada, preservando ao mesmo tempo a capacidade de uma sobreposição e soma a reconstrução com cancelamento do *aliasing* no domínio do tempo.

Resumindo, pode ser usado o contorno de pitch idealmente determinado sem a exigência de modificações adicionais no contorno de pitch, enquanto ao mesmo tempo permitindo uma representação dos blocos de entrada amostrados, que podem ser codificados de maneira eficiente usando uma subsequente transformada no domínio da frequência.

Breve Descrição dos Desenhos

São descritas na sequência várias configurações da presente invenção com referência às Figs. anexas, onde:

A Fig. 1 mostra uma configuração de um processador de áudio para a geração de uma representação processada de um sinal de áudio com uma sequência de frames;

5 As Figs. 2a a 2d mostram um exemplo de uma amostragem de um sinal de áudio de entrada que depende do contorno de pitch do sinal de áudio de entrada usando janelas de escalagem dependendo da amostragem aplicada;

10 A Fig. 3 mostra um exemplo de como associar as posições de amostragem usadas para amostragem e posições de amostragem de um sinal de entrada com amostras equidistantes;

A Fig. 4 mostra um exemplo de um contorno de tempo usado para determinar posições de amostragem de uma amostragem;

15 A Fig. 5 mostra uma configuração de uma janela de escalagem;

A Fig. 6 mostra uma exemplo de um contorno de pitch associado a uma sequência de frames de áudio para processamento;

20 A Fig. 7 mostra uma janela de escalagem aplicada a um bloco amostrado de transformada;

A Fig. 8 mostra as janelas de escalagem que correspondem ao contorno de pitch da Fig. 6;

25 A Fig. 9 mostra um outro exemplo de um contorno de pitch de uma sequência de frames de um sinal de áudio a ser processado;

A Fig. 10 mostra as janelas de escalagem usadas para o contorno de pitch da Fig. 9;

A Fig. 11 mostra as janelas de escalagem da Fig.

10 transformada para escala linear de tempo;

A Fig. 11a mostra um outro exemplo de um contorno de pitch de uma sequência de frames;

5 A Fig. 11b mostra as janelas de escalagem que correspondem à Fig. 11a em uma escala linear de tempo;

A Fig. 12 mostra uma configuração de um método para a geração de uma representação processada de um sinal de áudio;

10 A Fig. 13 mostra uma configuração de um processador para o processamento de representações amostradas de um sinal de áudio composto de uma sequência de frames de áudio; e

A Fig. 14 mostra uma configuração de um método para o processamento de representações amostradas de um sinal de áudio.

15 Descrição detalhada das configurações preferidas

A Fig. 1 mostra uma configuração de um processador de áudio 10 (sinal de entrada) para a geração de uma representação processada de um sinal de áudio tendo uma sequência de frames. O processador de áudio 2 compreende um amostrador 4, adaptado para amostrar uma entrada de sinal de áudio 10 (sinal de entrada) no processador de áudio 2 de maneira a obter os blocos de sinais (representações amostradas) usados como base para uma transformada no domínio da frequência. O processador de áudio 2 ainda compreende uma calculadora de janela transformada 6 adaptada para obter janelas de escalagem para as representações amostradas que saem do amostrador 4. Estas são enviadas a um janelador 8, adaptado para aplicar as janelas de escalagem às representações amostradas obtidas pelo amostrador 4. Em certas configurações, o

20

25

janelador pode também compreender um transformador no domínio da frequência 8a para obter representações no domínio da frequência das representações amostradas escaladas. Estas podem então ser processadas ou ainda transmitidas como uma representação codificada do sinal de áudio 10. O processador de áudio ainda usa um contorno de pitch 12 do sinal de áudio, que pode ser enviado ao processador de áudio ou que pode, de acordo com outra configuração, ser obtido pelo processador de áudio 2. Portanto, o processador de áudio 2 pode compreender opcionalmente um estimador de pitch para a obtenção do contorno de pitch.

O amostrador 4 pode operar em um sinal de áudio contínuo ou, de forma alternativa, em uma representação preamostrada do sinal de áudio. Neste ultimo caso, o amostrador pode reamostrar o sinal de áudio provido em sua entrada como indicado nas Figs. 2a a 2d. O amostrador esta adaptado para amostrar blocos de áudio sobrepostos vizinhos de maneira que a poção sobreposta tem o mesmo ou similar contorno de pitch dentro de cada um dos blocos de entrada após a amostragem.

O caso de um sinal preamostrado de áudio e elaborado em maiores detalhes na descrição das Figs. 3 e 4.

A calculadora de janela transformada 6 obtém as janelas de escalagem para os blocos de áudio dependendo da reamostragem feita pelo amostrador 4. Para tanto, pode existir um bloco de ajuste da taxa opcional de amostragem 14 para definir a regra de reamostragem usada pelo amostrador, que é então também enviada para a calculadora de janela transformada. Em uma configuração alternativa, o bloco de ajuste da taxa de amostragem 14 pode ser omitido e o contorno de pitch 12 pode ser diretamente

enviado à calculadora de janela transformada 6, que pode por si fazer os cálculos adequados. Além disso, o amostrador 4 pode comunicar a amostragem aplicada à calculadora de janela transformada 6 para permitir o cálculo das janelas de escalagem 5 adequadas.

A reamostragem é feita de maneira que o contorno de pitch dos blocos amostrados de áudio amostrados pelo amostrador 4 seja mais constante que o contorno de pitch do sinal original de áudio dentro do bloco de entrada. Para tanto, é avaliado o 10 contorno de pitch, como indicado para um exemplo específico nas Figs. 2a e 2d.

A Fig. 2a mostra um contorno de pitch linearmente degradado como uma função dos números de amostras do sinal preamostrado de entrada de áudio. Isto é, as Figs. 2a a 2d 15 ilustram um cenário em que os sinais de entrada de áudio já são providos como valores de amostras. Não obstante, os sinais de áudio antes da reamostragem e depois da reamostragem (warping da escala de tempo) também são ilustrados como sinais contínuos para ilustrar mais claramente o conceito. A Fig. 2b mostra uma exemplo 20 de um sinal seno 16 tendo uma frequência de varredura caindo das maiores frequências para menores frequências. Este comportamento corresponde ao contorno de pitch da Fig. 2a, que é mostrado em unidades arbitrárias. Novamente, é indicado que o warping de tempo do eixo do tempo é equivalente a uma reamostragem do sinal com 25 intervalos de amostragem de adaptação local.

Para ilustrar a sobreposição e somar processamento, a Fig. 2b mostra três frames consecutivos 20a, 20b e 20c do sinal de áudio, que são processados em bloco com uma

sobreposição de um frame (frame 20b). Isto é, um primeiro bloco de sinais 22 (bloco de sinais 1) que compreende as amostras do primeiro frame 20a e do segundo frame 20b é processado e reamostrado e um segundo bloco de sinais 24 compreendendo as amostras do segundo frame 20b e do terceiro frame 20c é reamostrado de forma independente. O primeiro bloco de sinais 22 é reamostrado para se obter a primeira representação reamostrada 26 mostrada na Fig. 2c e o segundo bloco de sinais 24 é reamostrado na segunda representação reamostrada 28 mostrada na Fig. 2d. Entretanto, a amostragem é feita de maneira que as porções correspondentes ao frame de sobreposição 20b têm o mesmo contorno de pitch ou somente um contorno de pitch pouco desviado (dentro de uma faixa pré-determinada de tolerância idêntica) na primeira representação amostrada 26 e na segunda representação amostrada 28. Claro, isto é somente verdade quando o pitch é estimado em termos de números de amostras. O primeiro bloco de sinais 22 é reamostrado para a primeira representação reamostrada 26, tendo um (idealizado) pitch constante. Assim, usando os valores de amostra da representação reamostrada 26 como entrada para uma transformada no domínio da frequência, idealmente somente um único coeficiente de frequência seria obtido. Isto é evidentemente uma representação extremamente eficiente do sinal de áudio. Os detalhes sobre como fazer a reamostragem é feita serão discutidos a seguir com referência às Figs. 3 e 4. Como se torna aparente a partir da Fig. 2c, a reamostragem é feita de maneira que o eixo das posições de amostra (o eixo x), que corresponde ao eixo do tempo em uma representação amostrada de forma equidistante é modificada de maneira que a forma do sinal resultante tenha somente uma única

frequência de pitch. Isto corresponde a um warping de tempo do eixo dos x e a uma subsequente amostragem equidistante da representação do warp de tempo do sinal do primeiro bloco de sinais 22.

5 O segundo bloco de sinais 24 é reamostrado de maneira que a porção do sinal que corresponde ao frame de sobreposição 20b na segunda representação reamostrada 28 tenha um contorno de pitch idêntico ou com pequeno desvio com relação à correspondente porção de sinal da representação reamostrada 26.

10 Entretanto, as taxas de amostragem diferem. Isto é, idênticas formas de sinais dentro das representações reamostradas são representadas por diferentes números de amostras. Não obstante, cada representação reamostrada, quando codificada por um codificador de transformada, resulta em uma representação

15 codificada altamente eficiente tendo somente um número limitado de coeficientes de frequência não zero.

Devido à reamostragem, a porção de sinais da primeira metade do bloco de sinais 22 é mudada para as amostras que pertencem à segunda metade do bloco de sinais da representação

20 reamostrada, como indicado na Fig. 2c. Em particular, a área hachurada 30 e o correspondente sinal à direita do segundo pico (indicado por II) é mudado para a metade direita da representação reamostrada 26 sendo, portanto, representado pela segunda metade das amostras da representação reamostrada 26. Entretanto, essas

25 amostras não têm porção correspondente de sinal na metade esquerda da representação reamostrada 28 da Fig. 2d.

Em outras palavras, apesar da reamostragem, a taxa de amostragem é determinada para cada bloco MDCT, de maneira

que a taxa de amostragem leva à uma duração constante no tempo linear do centro do bloco, que contém N amostras no caso de uma resolução de frequência de N e um comprimento máximo de janela de $2N$. No exemplo anteriormente descrito da Figs. 2a a 2d, $N = 1024$ e, como consequência, $2N = 2048$ amostras. A reamostragem realiza a interpolação real do sinal na posições exigidas. Devido à sobreposição dos dois blocos, que podem ter diferentes taxas de amostragem, a reamostragem deve ser feita duas vezes para cada segmento de tempo (equalizando um dos frames 20a a 20c) do sinal de entrada. O mesmo contorno de pitch, que controla o codificador ou o processador de áudio que realize a codificação, podem ser usados para controlar o processamento necessário para inverter a transformada e o warping, já que pode ser implementado dentro de um decodificador de áudio. Portanto, em certas configurações, o contorno de pitch é transmitido como informações auxiliares. Para evitar um descompasso entre um codificador e um correspondente decodificador, algumas configurações de codificadores usam o contorno de pitch codificado e, depois decodificado ao invés do contorno de pitch obtido originalmente ou entrado. Entretanto, o contorno de pitch obtido ou entrado pode, de forma alternativa, ser usado diretamente.

Para garantir que somente as correspondentes porções de sinais sejam sobrepostas na sobreposição e somem a reconstrução, são obtidas janelas adequadas de escalagem. Essas janelas de escalagem devem ser responsáveis pelo efeito que diferentes porções de sinais dos sinais originais sejam representadas dentro das correspondentes metades de janelas das representações reamostradas, por ser causada pela reamostragem

anteriormente descrita.

Janelas adequadas de escalagem podem ser obtidas para os sinais a codificar, que dependem de uma amostragem ou reamostragem aplicada para obter uma primeira e uma segunda 5 representações amostradas 26 e 28. Para o exemplo do sinal original ilustrado na Fig. 2b e o contorno de pitch ilustrado na Fig. 2a, são fornecidas janelas adequadas de escalagem para a segunda metade de janela da primeira representação amostrada 26 e para a primeira metade de janela da segunda representação 10 amostrada 28 pela primeira janela de escalagem 32 (sua segunda metade) e pela segunda janela de escalagem 34, respectivamente (a metade esquerda da janela correspondente às primeiras 1024 amostras da segunda representação amostrada 28).

Quando a porção de sinal dentro da área hachurada 15 30 da primeira representação amostrada 26 não tiver a correspondente porção de sinal na primeira metade da janela da segunda representação amostrada 28, a porção de sinal dentro da área hachurada deve ser totalmente reconstruída pela primeira representação amostrada 26. Em uma reconstrução MDCT, isto pode 20 ser feito quando as amostras correspondentes não são usadas para o *fading in* ou *fading out*, isto é, quando as amostras recebem um fator de escalagem de 1. Portanto, as amostras da janela de escalagem 32 correspondentes à área hachurada 30, são ajustadas à unidade. Ao mesmo tempo, o mesmo número de amostras deve ser 25 estabelecido em 0 no final da janela de escalagem para evitar uma mistura daquelas amostras com as amostras da primeira área sombreada 30 devido às propriedades inerentes da transformada MDCT e da transformada inversa.

Devido à reamostragem (aplicada), que alcança um idêntico warping de tempo do segmento de janela de sobreposição, aquelas amostras da segunda área sombreada 36 também não têm contraparte de sinal dentro da primeira metade da janela da segunda representação amostrada 28. Assim, esta porção de sinal pode ser totalmente reconstruída pela segunda metade de janela da segunda representação amostrada 28. O ajuste das amostras da primeira janela de escalagem correspondentes à segunda área sombreada 36 em 0 é, portanto possível sem perder as informações no sinal a ser reconstruído. Cada porção de sinal presente na primeira metade de janela da segunda representação amostrada 28 tem uma contraparte correspondente dentro da segunda metade de janela da primeira representação amostrada 26. Portanto, todas as amostras dentro da primeira metade de janela da segunda representação amostrada 28 são usadas para o *cross-fade* entre a primeira e a segunda representações amostradas 26 e 28, como indicado pela forma da segunda janela de escalagem 34.

Resumindo, a reamostragem dependente de pitch e o uso de janelas de escalagem projetadas de forma adequada permitem aplicar um contorno de pitch ideal, que não precisa obedecer a qualquer restrição além de ser contínuo. Como para efeito do aumento de eficiência de codificação, somente são relevantes as alterações relativas de pitch, o contorno de pitch pode ser mantido constante dentro e nos limites dos intervalos dos sinais onde não possam ser estimados pitches distintos ou quando não existir variações de pitch. Alguns conceitos alternativos propõem implementar o warping de tempo com contornos especializados de pitch ou de funções de warping de tempo, que possuem restrições

especiais com relação a seus contornos. Usando configurações da invenção, a eficiência de codificação será maior, já que o contorno ideal de pitch pode ser usado a qualquer momento.

Com relação às Figs. 3 a 5, uma determinada
5 possibilidade para a realização da reamostragem e para obter as janelas associadas de escalagem serão agora descritas em maiores detalhes.

Novamente a amostragem se baseia em um contorno de pitch linearmente decrescente 50, correspondente a um número
10 predeterminado de amostras N. O sinal correspondente 52 é ilustrado em tempo normalizado. No exemplo escolhido, o sinal tem a duração de 10 milissegundos. Se um sinal preamostrado é processado, o sinal 52 é amostrado normalmente em intervalos equidistantes de amostragem, como indicado pelas marcas de
15 conferência no eixo do tempo 54. Caso seja aplicado warping de tempo transformando de forma adequada o eixo dos tempos 54, o sinal 52 se tornaria, em uma escala de tempo em warp 56, um sinal 58, que tem um constante pitch. Isto é, a diferença de tempo (a diferença dos números de amostras) do sinal máximo vizinho 58 é
20 igual na nova escala de tempo 56. O comprimento do frame de sinal também mudaria para um novo comprimento de x milissegundos, dependendo do warping aplicado. Deve ser notado que o quadro de warping de tempo é somente usado para visualizar a idéia de reamostragem não equidistante usada em várias configurações da
25 presente invenção, que podem na verdade, ser implementadas somente usando os valores do contorno de pitch 50.

A seguinte configuração, que descreve como a amostragem pode ser feita, para a facilidade de compreensão,

baseada na suposição que o pitch alvo para o qual o sinal será dobrado (um pitch obtido de uma representação amostrada ou reamostrada do sinal original), é a unidade. Entretanto, não é necessário dizer que as seguintes considerações podem se aplicar
5 facilmente a pitches alvo arbitrários dos segmentos processados de sinal.

Supondo que o warping de tempo seria aplicado a um frame j iniciando na amostra jN , de modo a forçar o pitch para a unidade (1), a duração de frame após o warping de tempo
10 corresponderia a soma das N amostras correspondentes do contorno de pitch:

$$D_j = \sum_{i=0}^{N-1} \text{pitch_contour}_{jN+i}$$

Isto é, a duração do sinal de tempo em warp 58 (o tempo $t' = x$ na Fig. 3) é determinada pela fórmula acima.

15 Para obter N amostras em warp, o intervalo de amostragem no frame de tempo em warp j deve ser igual a:

$$I_j = N / D_j$$

Um contorno de tempo, que associa as posições das amostras originais com relação à janela MDCT em warp, pode ser
20 iterativamente construído de acordo com:

$$\text{time_contour}_{i+1} = \text{time_contour}_i + \text{pitch_contour}_{jN+i} * I_j$$

Um exemplo de um contorno de tempo é dado na Fig. 4. O eixo dos x mostra o número de amostras da representação reamostrada e o eixo dos y dá a posição deste número de amostragem
25 em unidades de amostras da representação original. Portanto, no exemplo da Fig. 3, o contorno de tempo é construído em etapas de tamanhos sempre decrescentes. A posição da amostra associada ao

número de amostras 1 na representação de warp de tempo (eixo n') em unidades da amostras originais é, por exemplo, aproximadamente 2. Para reamostragem não equidistante, dependente de contorno de pitch, as posições das amostras de entrada MDCT em warp são necessárias em unidades da escala de tempo original sem warp. A posição da amostra em warp da entrada MDCT i (eixo dos y) pode ser obtida procurando um par de posições originais de amostras k e $k+1$, que definem um intervalo incluindo i :

$$time_contour_k \leq i < time_contour_{k+1}.$$

Por exemplo, a amostra $i=1$ está localizada no intervalo definido pela amostra $k=0$, $k+1=1$. Uma parte fracionária u da posição da amostra é obtida supondo um contorno linear de tempo entre $k=0$ e $k+1=1$ (eixo dos x). Em termos gerais, a parte fracionária u da amostra i é determinada por:

$$u = \frac{i - time_contour_k}{time_contour_{k+1} - time_contour_k}.$$

Assim, a posição de amostragem da reamostragem não equidistante do sinal original pode ser obtida em unidades de posições originais de amostragem. Portanto, o sinal pode ser reamostrado de maneira que os valores reamostrados correspondam a um sinal de tempo em warp. Esta reamostragem pode, por exemplo, ser implementada usando um filtro de interpolação polifase h dividido em P subfiltros h_p com uma precisão de $1/P$ intervalos originais de amostras. Para tanto, o índice de subfiltro pode ser obtido a partir da posição fracional da amostra:

$$p = \lfloor uP \rfloor,$$

e a amostra em warp de entrada MDCT xw_i pode então ser calculada por convolução:

$$xw_i = x_k * h_{p,k}.$$

É claro que podem ser usados outros métodos de reamostragem como, por exemplo, reamostragem por splines, interpolação linear, interpolação quadrática ou outros métodos de reamostragem.

Após ter obtido as representações reamostradas, são obtidas as adequadas janelas de escalagem de maneira que nenhuma das duas janelas de sobreposição variem mais que $N/2$ amostras na área central do frame MDCT vizinho. Como anteriormente descrito, isto pode ser feito usando o contorno de pitch ou os correspondentes intervalos de amostras I_j ou, de forma equivalente, as durações de frame D_j . O comprimento de um overlap de frame "esquerdo" j (isto é, o *fade-in* com relação ao frame anterior $j-1$) é determinado por:

$$\sigma_l = \begin{cases} N/2 & \text{if } D_j \leq D_{j-1} \\ N/2 * D_{j-1} / D_j & \text{else} \end{cases},$$

[if = se] [else = ou]

e o comprimento do overlap "direito" do frame j (isto é, o *fade-out* do frame $j+1$ subsequente) é determinado por:

$$\sigma_r = \begin{cases} N/2 & \text{if } D_j \leq D_{j+1} \\ N/2 * D_{j+1} / D_j & \text{else} \end{cases}.$$

Assim, uma janela resultante para o frame j de comprimento $2N$, isto é, o comprimento típico de janela MDCT usado para reamostragem de frames com N amostras (isto é, a resolução de frequência de N), consiste dos seguintes segmentos, como ilustrados na Fig. 5.

$0 \leq i < N/2 - \sigma_l$	0
-----------------------------	---

$N/2 - \sigma l_j \leq i < N/2 + \sigma l_j$	$w_l(i)$
$N/2 + \sigma l_j \leq i < 3N/2 + \sigma r_j$	1
$3N/2 - \sigma r_j \leq i < 3N/2 + \sigma r_j$	$w_r(i)$
$3N/2 + \sigma r_j \leq i < 2N$	0

Isto é, as amostras 0 a $N/2 - \sigma l_j$ do bloco de entrada j são 0 quando D_{j+1} é maior ou igual a D_j . As amostras no intervalo $[N/2 - \sigma l_j; N/2 + \sigma l_j]$ são usadas para fazer o *fade in* na janela de escalagem. As amostras no intervalo $[N/2 + \sigma l_j; N]$ são ajustadas na unidade. A metade de janela direita, isto é, a metade de janela usada para fazer o *fade out* das $2N$ amostras compreende um intervalo $[N; 3/2N - \sigma r_j]$, que é ajustado na unidade. As amostras usadas para fazer o *fade out* da janela estão contidas dentro do intervalo $[3/2N - \sigma r_j; 3/2N + \sigma r_j]$. As amostras no intervalo $[3/2N + \sigma r_j; 2N]$ são ajustadas em 0. Em termos gerais, são obtidas as janelas de escalagem que tenham números idênticos de amostras, em que o primeiro número de amostras usadas para fazer o *fade out* da janela de escalagem difere de um segundo número de amostras usadas para fazer o *fade in* da janela de escalagem.

A forma exata ou os valores de amostra que correspondem às janelas de escalagem obtidas pode, por exemplo, ser obtida (também para um comprimento de overlap não inteiro) de uma interpolação linear das metades da janela protótipo, que especificam a função de janela nas posições de amostra inteira (ou em uma grade fixa com até maior resolução temporal). Isto é, as janelas protótipo são escaladas no tempo de acordo com os necessários comprimentos de *fade-in* e *fade-out* de $2\sigma l_j$ ou $2\sigma r_j$, respectivamente.

De acordo com outra configuração da presente

invenção, a porção da janela *fade-out* pode ser determinada sem usar informações sobre o contorno de pitch do terceiro frame. Para tanto, o valor de D_{j+1} pode ser limitado de maneira predeterminada.

Em certas configurações, o valor pode ser estabelecido em um número fixo predeterminado e a porção de janela *fade-in* do segundo bloco de entrada pode ser calculada com base na amostragem aplicada para obter a primeira representação amostrada, a segunda representação amostrada e o número predeterminado ou o limite predeterminado de D_{j+1} . Isto pode ser usado em aplicações em que

pequenos tempos de retardo sejam

de grande importância, já que cada bloco de entrada pode ser processado sem o conhecimento do bloco subsequente.

em outra configuração da presente invenção, o comprimento variável das janelas de escalagem pode ser usado para alternar entre os blocos de entrada de comprimentos diferentes.

As Figs. 6 a 8 ilustram um exemplo com uma resolução de frequência de $N=1024$ e um pitch de degradação linear. A Fig. 6 mostra o pitch como uma função do número de amostras.

Como se torna aparente, a degradação de pitch é linear e varia entre 3500 Hz e 2500 Hz no centro do bloco MDCT 1 (bloco de transformada 100), entre 2500 Hz e 1500 Hz no centro do bloco MDCT 2 (bloco de transformada 102) e entre 1500 Hz e 500 Hz no centro do bloco MDCT 3 (bloco de transformada 104). Isto corresponde às seguintes durações de frame na escala de tempo em warp (dada em unidades de duração (D_2) do bloco de transformada 102:

$$D_1 = 1.5D_2; D_3 = 0.5D_2.$$

Visto o acima, o segundo bloco de transformada

102 tem um comprimento de sobreposição esquerdo $\sigma_{l_2} = N/2 = 512$, já que $D_2 < D_1$ e um comprimento de sobreposição direito $\sigma_{r_2} = N/2 \times 0,5 = 256$. A Fig. 7 mostra a janela de escalagem calculada com as propriedades anteriormente descritas.

5 Além disso, o comprimento de sobreposição direito do bloco 1 é igual a $\sigma_{r_1} = N/2 \times 2/3 = 341,33$ e o comprimento de sobreposição esquerdo do bloco 3 (bloco de transformada 104) é $\sigma_{l_3} = N/2 = 512$. Como fica aparente, a forma das janelas de transformada somente depende do contorno de pitch do sinal
10 subjacente. A Fig. 8 mostra as janelas efetivas no domínio do tempo não dobrado (isto é, linear) dos blocos de transformadas 100, 102 e 104.

As Figs. 9 a 11 mostram outro exemplo de uma sequência de quatro bloco consecutivos de transformadas 110 a 113.
15 Entretanto, o contorno de pitch como indicado na Fig. 9 é um pouco mais complexo, tendo a forma de uma função senoidal. Para a resolução de frequência do exemplo $N(1024)$ e um comprimento máximo de janela 2048, as funções de janela adaptadas de acordo (calculadas) no domínio de tempo em warp são dadas na Fig. 10.
20 Suas formas efetivas correspondentes na escala linear de tempo estão ilustradas na Fig. 11. Pode ser notado que todas as Figs. mostram funções de janelas quadráticas para ilustrar as capacidades de reconstrução da sobreposição e o procedimento de soma melhor quando as janelas são aplicadas duas vezes (antes da MDCT e depois da IMDCT). O aliasing na propriedade de cancelamento
25 do domínio do tempo das janelas geradas pode ser reconhecido pelas simetrias das transições correspondentes no domínio em warp. Como anteriormente determinado, as Figs. também ilustram que menores

intervalos de transição podem ser selecionados nos blocos onde se reduz o pitch na direção dos limites, já que isto corresponde a aumentar os intervalos de amostragem e, portanto, das formas efetivas esticadas no domínio linear de tempo. Um exemplo deste comportamento pode ser visto no frame 4 (bloco de transformada 113), onde a função da janela se prolonga menos que o máximo de 2048 amostras. Entretanto, devido aos intervalos de amostragem, que são inversamente proporcionais ao pitch do sinal, sendo a duração máxima possível coberta sob a restrição que somente duas janelas sucessivas podem se sobrepor em qualquer ponto do tempo.

As Figs. 11a e 11b mostram outro exemplo de um contorno de pitch (informações de contorno de pitch) e de suas correspondentes janelas de escalagem em uma escala linear de tempo.

A Fig. 11a dá o contorno de pitch 120, como função dos números de amostras, que são indicados no eixo dos x. Isto é, a Fig. 11a dá as informações de contorno de warp desses três blocos consecutivos de transformação 122, 124 e 126.

A Fig. 11b ilustra as correspondentes janelas de escalagem para cada um dos blocos de transformadas 122, 124 e 126 em uma escala linear de tempo. As janelas transformadas são calculadas dependendo da amostragem aplicada ao sinal correspondente às informações de contorno de pitch ilustradas na Fig. 11a. Essas janelas transformadas são retransformadas na escala linear de tempo, para prover a ilustração da Fig. 11b.

Em outras palavras, a Fig. 11b ilustra que as re-janelas de escalagem transformadas podem ultrapassar o limite de frame (linhas sólidas da Fig. 11b) quando dobradas de volta ou

retransformadas para a escala linear de tempo. Isto pode ser considerado no codificador fornecendo mais algumas amostras de entrada além dos limites dos frames. No decodificador, o buffer de saída pode ser suficientemente grande para guardar as amostras correspondentes. Uma forma alternativa de considerar o fato pode ser encurtar a faixa de sobreposição da janela e usar ao invés disso regiões de zeros e uns, de maneira que a parte não zero da janela não ultrapasse os limites do frame.

Além disso, como se torna aparente pela Fig. 11b, as interseções das janelas redobradas (os pontos de simetria de aliasing no domínio do tempo) não são alteradas pelo warping de tempo, já que permanecem nas posições "não-warped" 512, 3x512, 5x512, 7x512. Este também é o caso das janelas de escalagem correspondentes no domínio warped, já que também existem simétricas às posições dadas por um quarto e três quartos do comprimento do bloco de transformada.

Uma configuração de um método para a geração de uma representação processada de um sinal de áudio tendo uma sequência de frames pode ser caracterizada pelas etapas ilustradas na Fig. 12.

Em uma etapa de amostragem 200, o sinal de áudio é amostrado dentro de um primeiro e um segundo frame da sequência de frames, o segundo frame seguindo o primeiro frame, usando informações em um contorno de pitch do primeiro e do segundo frame para obter uma primeira representação amostrada e o sinal de áudio é amostrado dentro do segundo e um terceiro frame, o terceiro frame seguindo o segundo frame na sequência de frames, usando informações no contorno de pitch do segundo frame e informações em

um contorno de pitch do terceiro frame para obter uma segunda representação amostrada.

Em uma etapa de cálculo de janela transformada 202, a primeira janela de escalagem é obtida para a primeira representação amostrada e a segunda janela de escalagem é obtida para a segunda representação amostrada, em que as janelas de escalagem dependem da amostragem aplicada para obter a primeira e a segunda representações amostradas.

Em uma etapa de janelamento 204, a primeira janela de escalagem é aplicada na primeira representação amostrada e a segunda janela de escalagem é aplicada na segunda representação amostrada.

A Fig. 13 mostra uma configuração de um processador de áudio 290 para o processamento de uma primeira representação amostrada de um primeiro e um segundo frame de um sinal de áudio tendo uma sequência de frames em que o segundo frame segue o primeiro frame e para os demais processamentos uma segunda representação amostrada do segundo frame e de um terceiro frame seguindo o segundo frame na sequência de frames, compreendendo:

A calculadora de janela transformada 300 adaptada para obter uma primeira janela de escalagem para a primeira representação amostrada 301a usando informações em um contorno de pitch 302 do primeiro e do segundo frame e para obter uma segunda janela de escalagem para a segunda representação amostrada 301b usando informações em um contorno de pitch do segundo e do terceiro frame, onde as janelas de escalagem têm número idênticos de amostras e onde um primeiro número de amostras usado para fazer

o *fade out* da primeira janela de escalagem difere de um segundo número de amostras usado para fazer o *fade in* da segunda janela de escalagem;

o processador de áudio 290 ainda compreende um
5 janelador 306 adaptado para aplicar a primeira janela de escalagem à primeira representação amostrada e para aplicar a segunda janela de escalagem à segunda representação amostrada. O processador de áudio 290, além disso, compreende um reamostrador 308 adaptado para reamostrar uma primeira representação amostrada escalada para
10 obter uma primeira representação reamostrada usando as informações no contorno de pitch do primeiro e do segundo frame e para reamostrar a segunda representação amostrada escalada para obter uma segunda representação reamostrada, usando as informações no contorno de pitch do segundo e do terceiro frame, de maneira que
15 uma porção do primeiro representação reamostrada correspondendo ao segundo frame tem um contorno de pitch dentro de uma faixa pré-determinada de tolerância de um contorno de pitch da porção da segunda representação reamostrada correspondendo ao segundo frame. Para obter a janela de escalagem, a calculadora de janela
20 transformada 300 pode tanto receber o contorno de pitch 302 diretamente ou receber informações da reamostragem de um ajustador opcional de taxa de amostra 310, que recebe o contorno de pitch 302 e que obtém uma estratégia de reamostragem.

Em outra configuração da presente invenção, um
25 processador de áudio, e além disso compreende um somador opcional 320, que é adaptado para somar a porção da primeira representação reamostrada correspondendo ao segundo frame e uma porção da segunda representação reamostrada correspondendo ao segundo frame

para obter uma representação reconstruída do segundo frame do sinal de áudio como um sinal de saída 322. A primeira representação amostrada e a segunda representação amostrada poderiam, em uma configuração, ser providas como uma saída para o processador de áudio 290. Em outra configuração, o processador de áudio pode, opcionalmente, compreender um transformador no domínio da frequência inversa 330, que pode obter a primeira e a segunda representações amostradas das representações no domínio da frequência da primeira e da segunda representações amostradas providas à entrada do transformador no domínio da frequência inversa 330.

A Fig. 14 mostra uma configuração de um método para o processamento de uma primeira representação amostrada de um primeiro e um segundo frame de um sinal de áudio tendo uma sequência de frames em que o segundo frame segue o primeiro frame e para o processamento de uma segunda representação amostrada do segundo frame e de um terceiro frame seguindo o segundo frame na sequência de frames. Em uma etapa de criação de janela 400, uma primeira janela de escalagem é obtida para a primeira representação amostrada usando informações em um contorno de pitch do primeiro e do segundo frame e uma segunda janela de escalagem é obtida para a segunda representação amostrada usando informações em um contorno de pitch do segundo e do terceiro frames, em que as janelas de escalagem têm números idênticos de amostras e onde um primeiro número de amostras usado para fazer o *fade out* da primeira janela de escalagem difere de um segundo número de amostras usadas para fazer o *fade in* da segunda janela de escalagem.

Em uma etapa de escalagem 402, a primeira janela de escalagem é aplicada à primeira representação amostrada e a segunda janela de escalagem é aplicada à segunda representação amostrada.

5 Em uma operação de reamostragem 402, a primeira representação amostrada escalada é reamostrada para obter uma primeira representação reamostrada usando as informações no contorno de pitch do primeiro e do segundo frames e a segunda representação amostrada escalada é reamostrada para obter uma
10 segunda representação reamostrada usando as informações no contorno de pitch do segundo e do terceiro frames, de maneira que uma porção da primeira representação reamostrada correspondendo ao primeiro frame tem um contorno de pitch dentro de uma faixa pré-determinada de tolerância de um contorno de pitch da porção da
15 segunda representação reamostrada correspondendo ao segundo frame.

De acordo com outra configuração da invenção, o método compreende uma etapa opcional de síntese 406 onde a porção da primeira representação reamostrada correspondendo ao segundo frame e a porção da segunda representação reamostrada
20 correspondendo ao segundo frame são combinadas para obter uma representação reconstruída do segundo frame do sinal de áudio.

Resumindo, as configurações previamente discutidas da presente invenção permitem a aplicação de um contorno ideal de pitch a um sinal de áudio contínuo ou
25 preamostrado para reamostrar ou transformar o sinal de áudio em uma representação, que pode ser codificada resultando em uma representação codificada com alta qualidade e baixa taxa de bits. Para que isto seja alcançado, o sinal reamostrado pode ser

codificado usando a transformada no domínio da frequência. Isto pode, por exemplo, ser a transformada co-seno discreta modificada discutida nas configurações anteriores. Entretanto, outras transformadas no domínio da frequência ou outras transformadas
5 podem, de forma alternativa, ser usadas para obter uma representação codificada de um sinal de áudio com baixa taxa de bits.

Não obstante, é também possível usar diferentes transformadas de frequência para se obter o mesmo resultado, como
10 por exemplo, uma transformada rápida de Fourier ou uma transformada discreta co-seno para obter a representação codificada do sinal de áudio.

Não é necessário dizer que o número de amostras, isto é, o bloco de transformadas usado como entrada para a
15 transformada no domínio da frequência não se limita ao exemplo particular usado nas configurações anteriormente descritas. Ao invés disso, pode ser usado um comprimento arbitrário de bloco de frame, como, por exemplo, blocos que consistem de 256, 512, 1024 blocos.

20 Podem ser usadas técnicas arbitrárias de amostragem ou de reamostragem dos sinais de áudio para implementar outras configurações da presente invenção.

Um processador de áudio usado para gerar a representação processada, como ilustrado na Fig. 1, pode receber o
25 sinal de áudio e as informações sobre o contorno de pitch como entradas separadas, por exemplo, como fluxos separados de bits de entrada. Em outras configurações, entretanto, o sinal de áudio e as informações sobre o contorno de pitch podem ser providos dentro

de um fluxos intercalados de bits, como as informações do sinal de áudio e o contorno de pitch são multiplexados pelo processador de áudio. As mesmas configurações podem ser implementadas para o processador de áudio, obtendo uma reconstrução do sinal de áudio com base nas representações amostradas. Isto é, as representações amostradas podem ser enviadas como um fluxo conjunto de bits junto com as informações de contorno de pitch ou como dois fluxos separados de bits. O processador de áudio poderia, além disso, compreender um transformador no domínio da frequência para transformar as representações reamostradas em coeficientes de transformadas, que são então transmitidos em conjunto com um contorno de pitch como uma representação codificada do sinal de áudio, de maneira a transmitir de maneira eficiente um sinal codificado de áudio a um decodificador correspondente.

As configurações anteriormente descritas, para o benefício da simplicidade, assumem que o pitch alvo para o qual o sinal reamostrado é a unidade. Não é preciso dizer que o pitch pode ser qualquer outro pitch arbitrário. Como o pitch pode ser aplicado sem restrições ao contorno de pitch, é também possível aplicar um contorno pitch constante no caso de nenhum contorno de pitch poder ser obtido ou no caso de nenhum contorno de pitch ser entregue.

Dependendo de determinadas exigências de implementação dos métodos do invento, os métodos do invento podem ser implementados em hardware ou em software. A implementação pode ser feita usando um meio de armazenamento digital, em particular um disco, DVD ou um CD tendo armazenado um controle de leitura de sinais eletrônicos, que coopera com um sistema de computador

programável, de maneira que os métodos do invento sejam realizados. Em geral, a presente invenção é, portanto, um produto de programa de computador com um código de programa armazenado em um veículo de leitura por máquina, o código de programa operando
5 para realizar os métodos do invento quando o produto de programa de computador opera em um computador. Em outras palavras, os métodos do invento são, portanto, um programa de computador tendo um código de programa para realizar pelo menos um dos métodos do invento quando o produto de programa de computador opera em um
10 computador.

Apesar de o citado ter sido apresentado e descrito em particular com referência a determinadas configurações em particular, será entendido pelos técnicos no assunto que várias outras alterações de forma e detalhes podem ser feitas sem
15 abandonar seu espírito e escopo. É compreendido que várias alterações podem ser feitas na adaptação a diferentes configurações sem abandonar os conceitos mais amplos revelados na presente e englobados pelas reivindicações a seguir.

REIVINDICAÇÕES

1. Processador de áudio para a geração de uma representação processada de um sinal de áudio dotado de uma sequência de frames, o processador de áudio caracterizado pelo fato de que compreende: um amostrador adaptado para amostrar o sinal de áudio dentro de um primeiro e um segundo frame da sequência de frames, o segundo frame seguindo o primeiro frame, o amostrador usando as informações sobre o contorno de pitch do primeiro e do segundo frame para obter uma primeira representação amostrada e para amostrar o sinal de áudio dentro do segundo e do terceiro frame, o terceiro frame seguindo o segundo frame na sequência de frames usando as informações sobre o contorno de pitch do segundo frame e as informações sobre o contorno de pitch do terceiro frame para obter uma segunda representação amostrada; uma calculadora de janela transformada adaptada para obter uma primeira janela de escalagem para a primeira representação amostrada e uma segunda janela de escalagem para a segunda representação amostrada, as janelas de escalagem dependendo da amostragem aplicada para obter a primeira representação amostrada ou a segunda representação amostrada; e um janelador adaptado para aplicar a primeira janela de escalagem à primeira representação amostrada e a segunda janela de escalagem à segunda representação amostrada para obter uma representação processada do primeiro, segundo e terceiro frames de áudio do sinal de áudio.

2. Processador de áudio, de acordo com a reivindicação 1, caracterizado pelo fato de que o amostrador opera para amostrar o sinal de áudio de maneira que um contorno de pitch dentro da primeira e da segunda representações amostradas seja

mais constante que o contorno de pitch do sinal de áudio dentro dos correspondentes primeiro, segundo e terceiro frames.

3. Processador de áudio, de acordo com a reivindicação 1, caracterizado pelo fato de que o amostrador opera para reamostrar um sinal de áudio amostrado com N amostras em cada primeiro, segundo e terceiro frames de maneira que cada primeira e segunda representações amostradas compreenda 2 N amostras.

4. Processador de áudio, de acordo com a reivindicação 3, caracterizado pelo fato de que o amostrador opera para obter uma amostra i da primeira representação amostrada na posição dada pela fração u entre as posições originais de amostragem k e (k+1) das 2N amostras do primeiro e do segundo frames, a fração u dependendo de um contorno de tempo que associa as posições de amostragem usadas pelo amostrador e as posições originais de amostragem do sinal de áudio amostrado do primeiro e do segundo frames.

5. Processador de áudio, de acordo com a reivindicação 4, caracterizado pelo fato de que o amostrador opera para usar um contorno de tempo obtido a partir do contorno de pitch p_i dos frames de acordo com a seguinte equação:

$$time_contour_{i+1} = time_contour_i + (p_i \times I),$$

onde um intervalo de referência de tempo I da primeira representação amostrada obtido a partir de um indicador de pitch D obtido a partir do contorno de pitch p_i de acordo com:

$$D = \sum_{i=0}^{2N-1} p_i, \quad I = 2N / D.$$

6. Processador de áudio, de acordo com a reivindicação 1, caracterizado pelo fato de que a calculadora de

janela transformada é adaptada para obter janelas de escalagem com números idênticos de amostras, onde um primeiro número de amostras usadas para fazer o *fade out* da primeira janela de escalagem difere de um segundo número de amostras usado para fazer o *fade in* da segunda janela de escalagem.

7. Processador de áudio, de acordo com a reivindicação 1, caracterizado pelo fato de que a calculadora de janela transformada é adaptada para obter uma primeira janela de escalagem em que um primeiro número de amostras é menor que um segundo número de amostras da segunda janela de escalagem quando os primeiro e segundo frames combinados têm maior pitch médio que o segundo e o terceiro frames combinados ou obtenham uma primeira janela de escalagem em que o primeiro número de amostras seja maior que o segundo número de amostras da segunda janela de escalagem quando o primeiro e o segundo frames combinados têm menor pitch médio que o segundo e terceiro frames combinados.

8. Processador de áudio, de acordo com a reivindicação 6, caracterizado pelo fato de que a calculadora de janela transformada é adaptada para obter janelas de escalagem onde um número de amostras antes de as amostras usadas para fazer o *fade out* e onde um número de amostras após as amostras usadas para fazer o *fade in* são ajustadas para a unidade e onde o número de amostras após as amostras para fazer o *fade out* e antes das amostras usadas para fazer o *fade in* são ajustadas em 0.

9. Processador de áudio, de acordo com a reivindicação 8, caracterizado pelo fato de que a calculadora de janela transformada é adaptada para obter o número de amostras usadas para fazer o *fade in* e usadas para fazer o *fade out*

dependente de um primeiro indicador de pitch D_j do primeiro e do segundo frames tendo amostras $0, \dots, 2N-1$ e de um segundo indicador de pitch D_{j+1} do segundo e do terceiro frames tendo amostras $N, \dots, 3N-1$, de maneira que o número de amostras usado para fazer o *fade in* seja:

$$N \text{ if } D_{j+1} \leq D_j \text{ or}$$

$$N \times \frac{D_j}{D_{j+1}} \text{ if } D_{j+1} > D_j; \text{ e}$$

o primeiro número de amostras usado para fazer o *fade out* seja:

$$N \text{ if } D_j \leq D_{j+1} \text{ or}$$

$$N \times \frac{D_{j+1}}{D_j} \text{ if } D_j > D_{j+1}$$

onde o indicador de pitches D_j e D_{j+1} seja obtido a partir do contorno de pitch p_i de acordo com as seguintes equações:

$$D_{j+1} = \sum_{i=N}^{3N-1} p_i \text{ and } D_j = \sum_{i=0}^{2N-1} p_i .$$

10. Processador de áudio, de acordo com a reivindicação 8, caracterizado pelo fato de que o calculador de janela opera para obter um primeiro e um segundo número de amostras pela reamostragem de uma janela predeterminada de *fade in* e *fade out* com iguais números de amostras para o primeiro e o segundo número de amostras.

11. Processador de áudio para o processamento de uma primeira representação amostrada de um primeiro e um segundo frame de um sinal de áudio tendo a sequência de frames em que o segundo frame segue o primeiro frame e para o processamento de uma

segunda representação amostrada do segundo frame e de um terceiro frame do sinal de áudio seguindo o segundo frame na sequência de frames, compreendendo:

uma calculadora de janela transformada adaptada para obter uma primeira janela de escalagem para a primeira representação amostrada usando informações em um contorno de pitch do primeiro e do segundo frames e para obter uma segunda janela de escalagem para a segunda representação amostrada usando informações em um contorno de pitch do segundo e do terceiro frames, caracterizado pelo fato de que as janelas de escalagem têm um número idêntico de amostras e onde um primeiro número de amostras usado para fazer o *fade out* da primeira janela de escalagem difere de um segundo número de amostras usado para fazer o *fade in* da segunda janela de escalagem; um janelador usado para aplicar a primeira janela de escalagem à primeira representação amostrada e para aplicar a segunda janela de escalagem à segunda representação amostrada;

e um reamostrador adaptado para reamostrar a primeira representação amostrada escalada para obter uma primeira representação reamostrada usando as informações no contorno de pitch do primeiro e do segundo frame e para reamostrar a segunda representação amostrada escalada para obter a segunda representação reamostrada usando as informações no contorno de pitch do segundo e do terceiro frames, a reamostragem dependendo das janelas de escalagem obtidas.

12. Processador de áudio, de acordo com a reivindicação 11, caracterizado pelo fato de que compreende um somador adaptado para somar a porção da primeira representação

reamostrada correspondente ao segundo frame e a porção da segunda representação reamostrada correspondente ao segundo frame para obter a representação reconstruída do segundo frame do sinal de áudio.

13. Método para a geração de uma representação processada de um sinal de áudio tendo uma sequência de frames caracterizado pelo fato de que compreende:

amostrar o sinal de áudio dentro de um primeiro e de um segundo frame da sequência de frames, o segundo frame seguindo o primeiro frame, a amostragem usando informações em um contorno de pitch do primeiro e do segundo frame para obter a primeira representação amostrada;

amostrar o sinal de áudio dentro do segundo e de um terceiro frame, o terceiro frame seguindo o segundo frame na sequência de frames, a amostragem usando as informações no contorno de pitch do segundo frame e informações em um contorno de pitch do terceiro frame para obter a segunda representação amostrada; obter uma primeira janela de escalagem para a primeira representação amostrada e uma segunda janela de escalagem para a segunda representação amostrada, as janelas de escalagem dependendo das amostragens aplicadas para obter a primeira representação amostrada ou a segunda representação amostrada; e

aplicar a primeira janela de escalagem à primeira representação amostrada e aplicar a segunda janela de escalagem à segunda representação amostrada.

14. Método para o processamento de uma primeira representação amostrada de um primeiro e um segundo frames de um sinal de áudio tendo uma sequência de frames onde o segundo frame

segue o primeiro frame e para processamento de uma segunda representação amostrada do segundo frame e de um terceiro frame do sinal de áudio seguindo o segundo frame na sequência de frames, caracterizado pelo fato de que compreende:

obter uma primeira janela de escalagem para a primeira representação amostrada usando informações em um contorno de pitch do primeiro e do segundo frames e obtendo uma segunda janela de escalagem para a segunda representação amostrada usando informações em um contorno de pitch do segundo e do terceiro frames,

em que as janelas de escalagem são obtidas de maneira que tenham um número idêntico de amostras, onde um primeiro número de amostras usado para fazer o *fade out* da primeira janela de escalagem difere de um segundo número de amostras usado para fazer o *fade in* da segunda janela de escalagem; aplicar a primeira janela de escalagem à primeira representação amostrada e a segunda janela de escalagem à segunda representação amostrada; e reamostrar a primeira representação amostrada escalada para obter uma primeira representação reamostrada usando as informações no contorno de pitch do primeiro e do segundo frames e reamostrar a segunda representação amostrada escalada para obter a segunda representação reamostrada usando as informações no contorno de pitch do segundo e do terceiro frames, a reamostragem dependendo das janelas de escalagem obtidas.

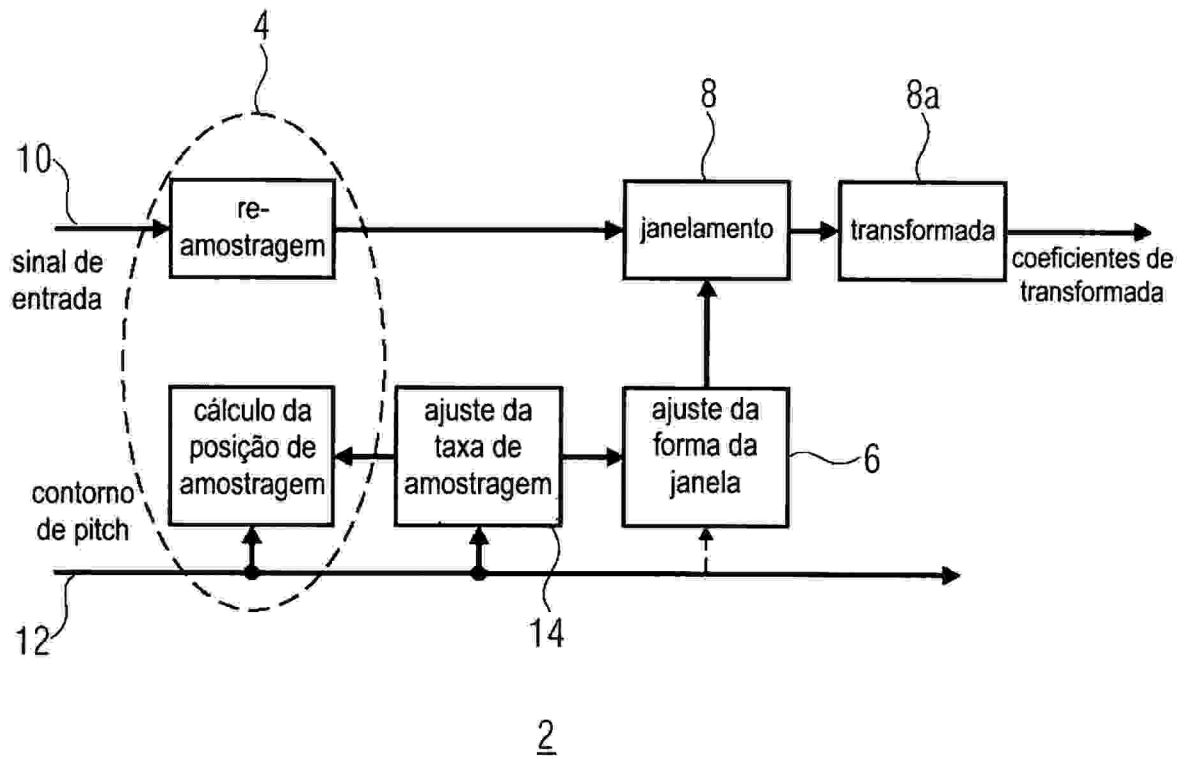
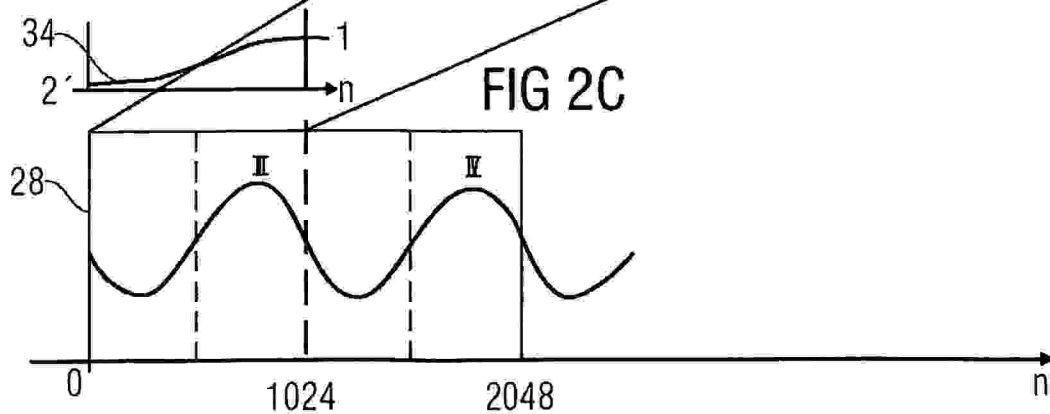
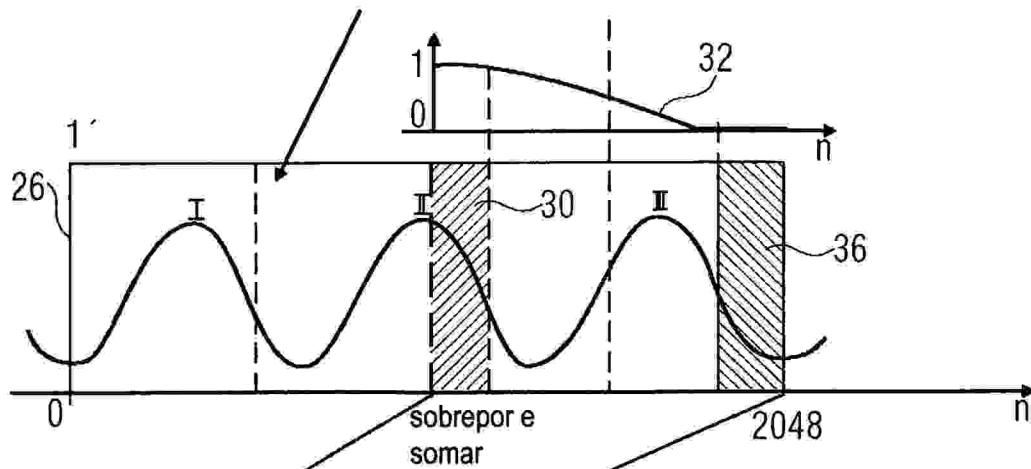
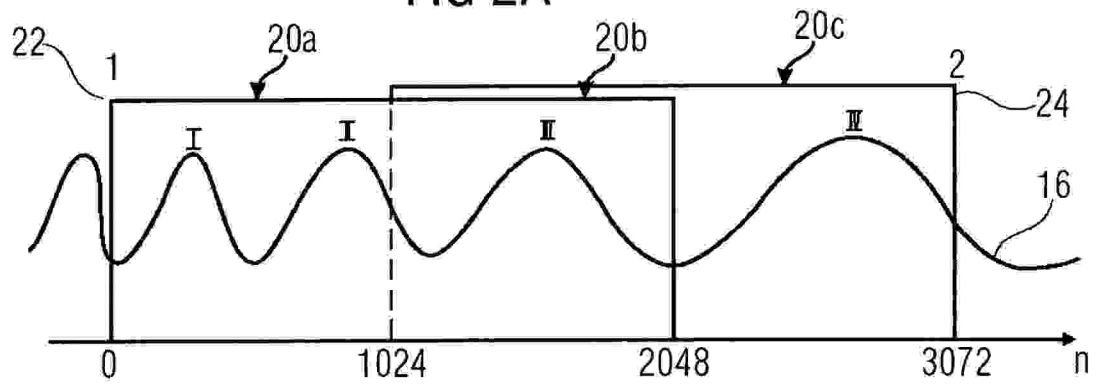
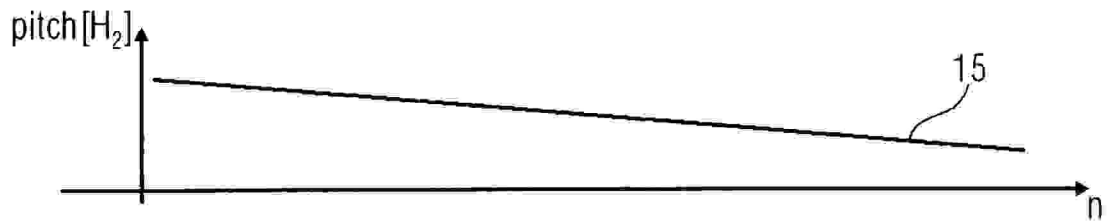


FIG 1

2/16



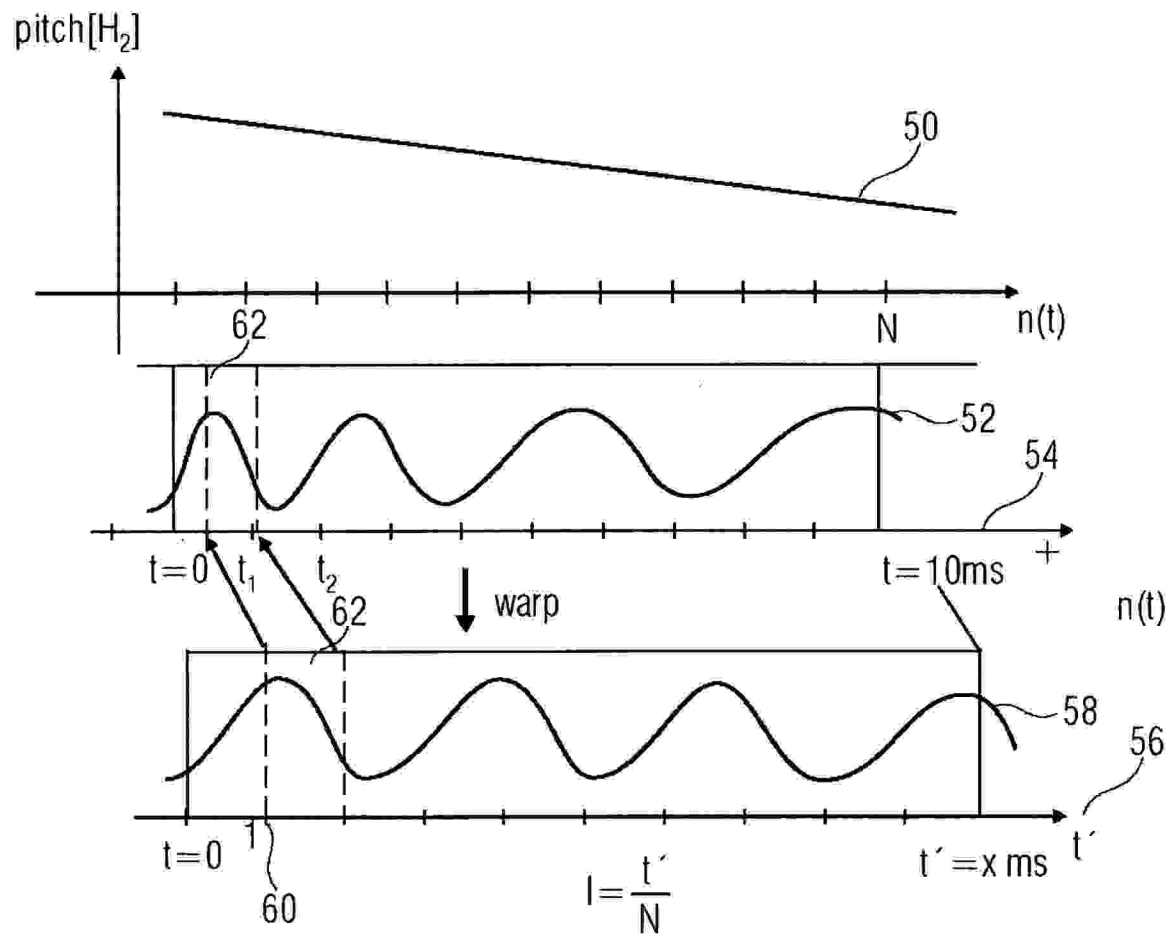


FIG 3

contorno de tempo

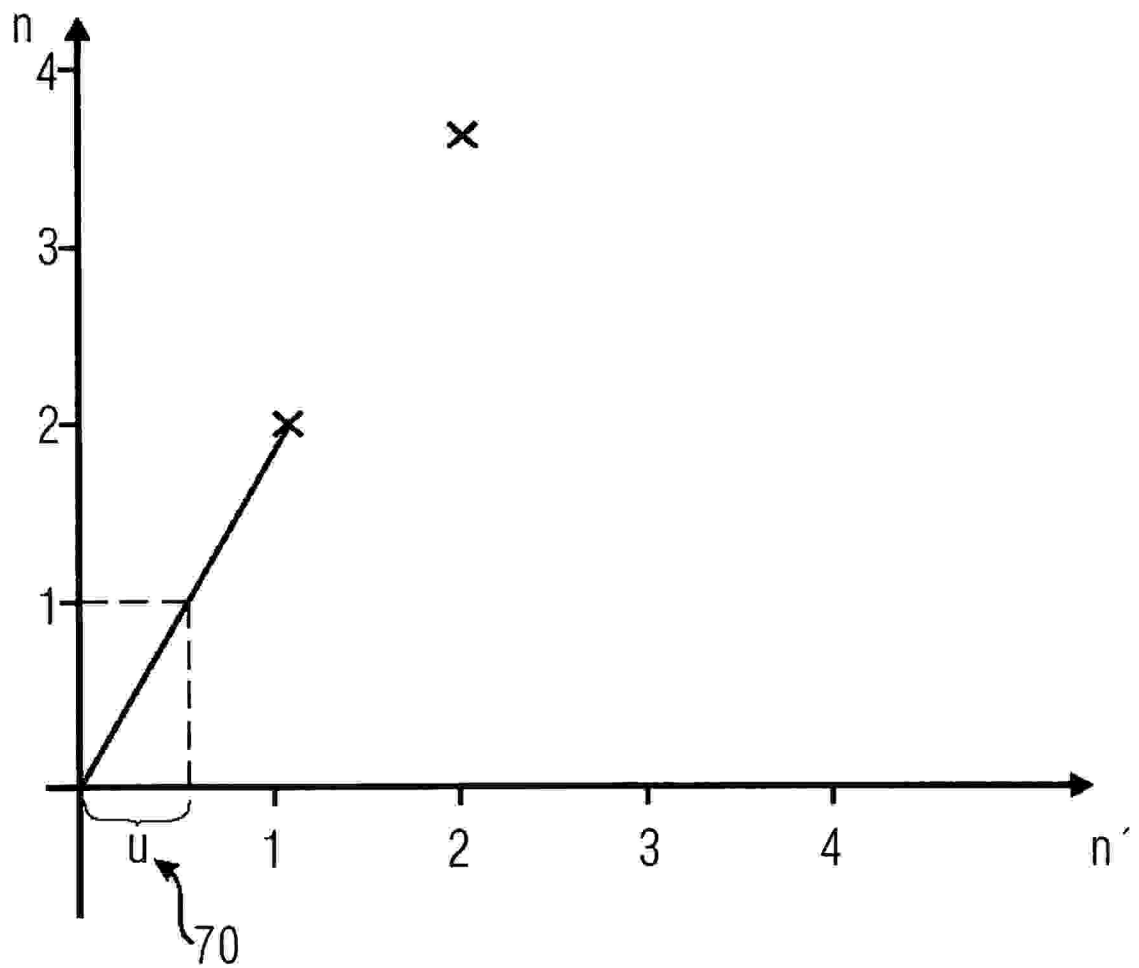


FIG 4

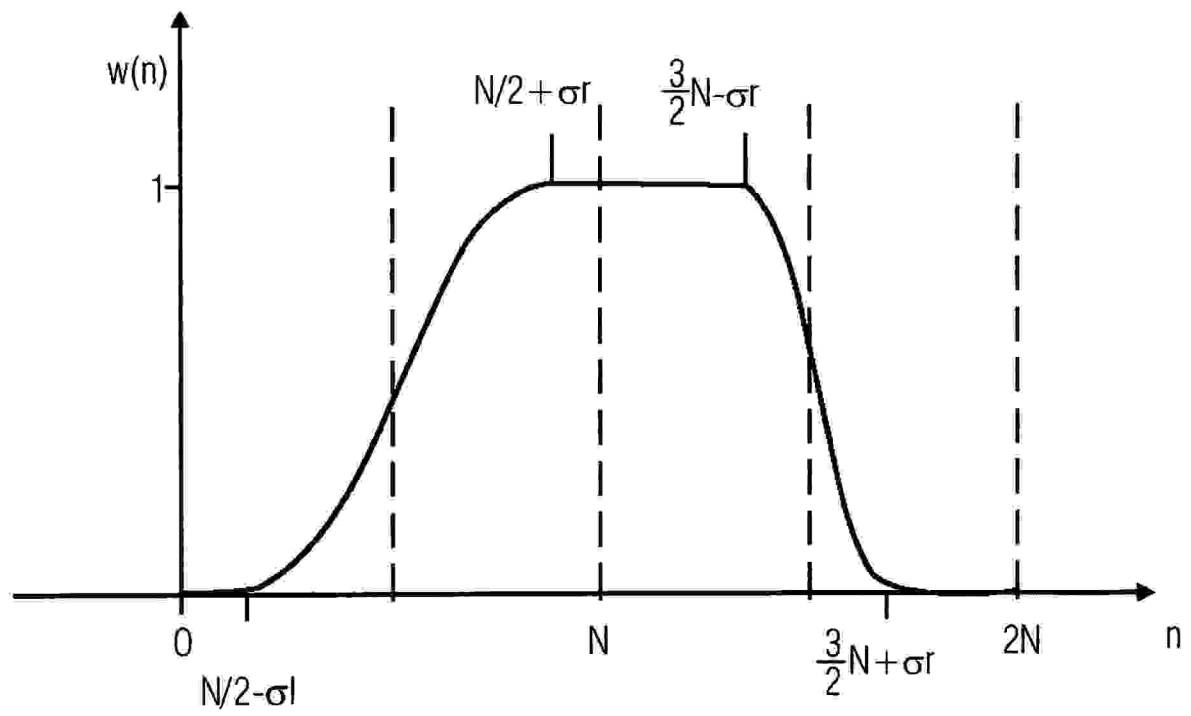


FIG 5

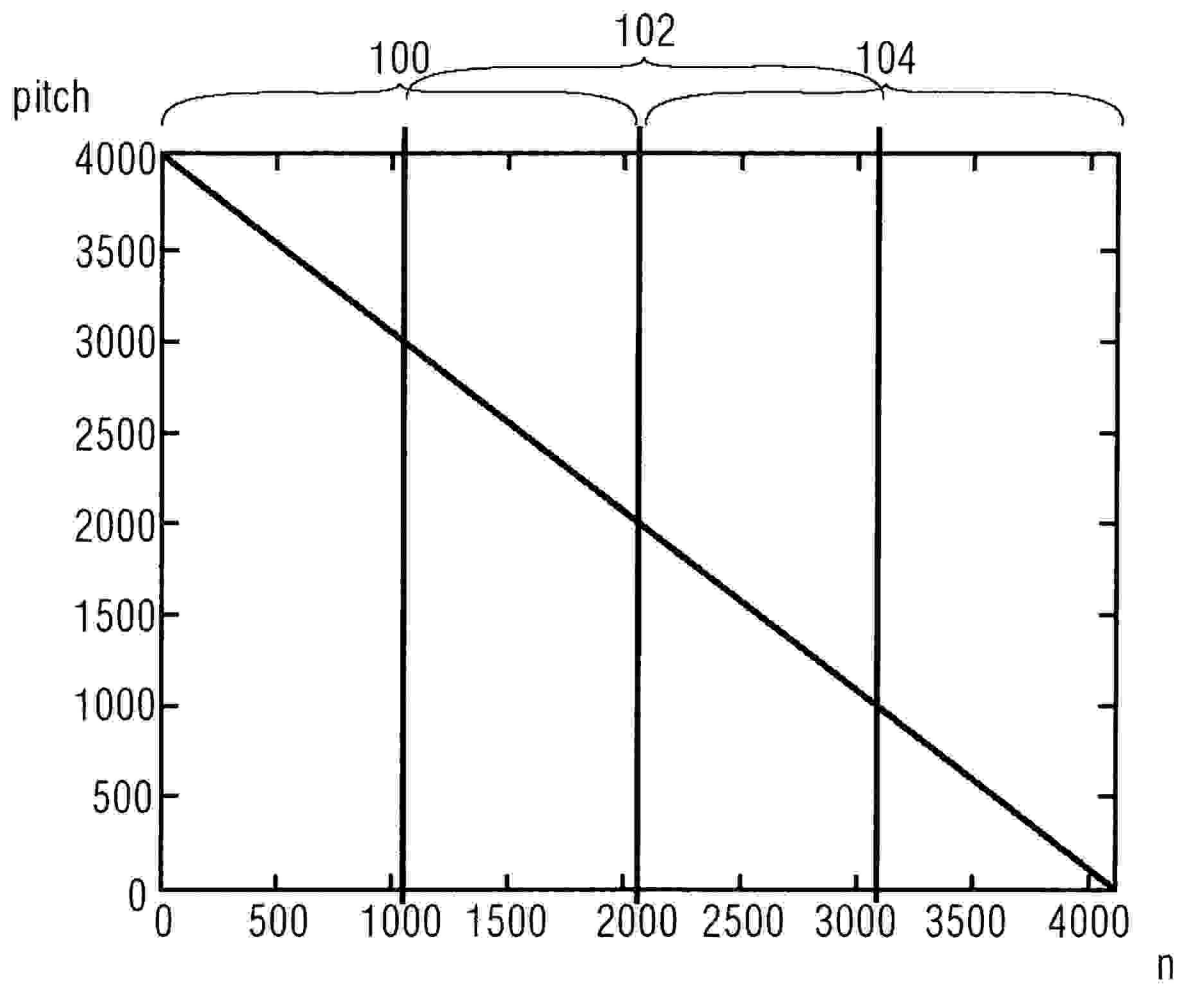


FIG 6

7/16

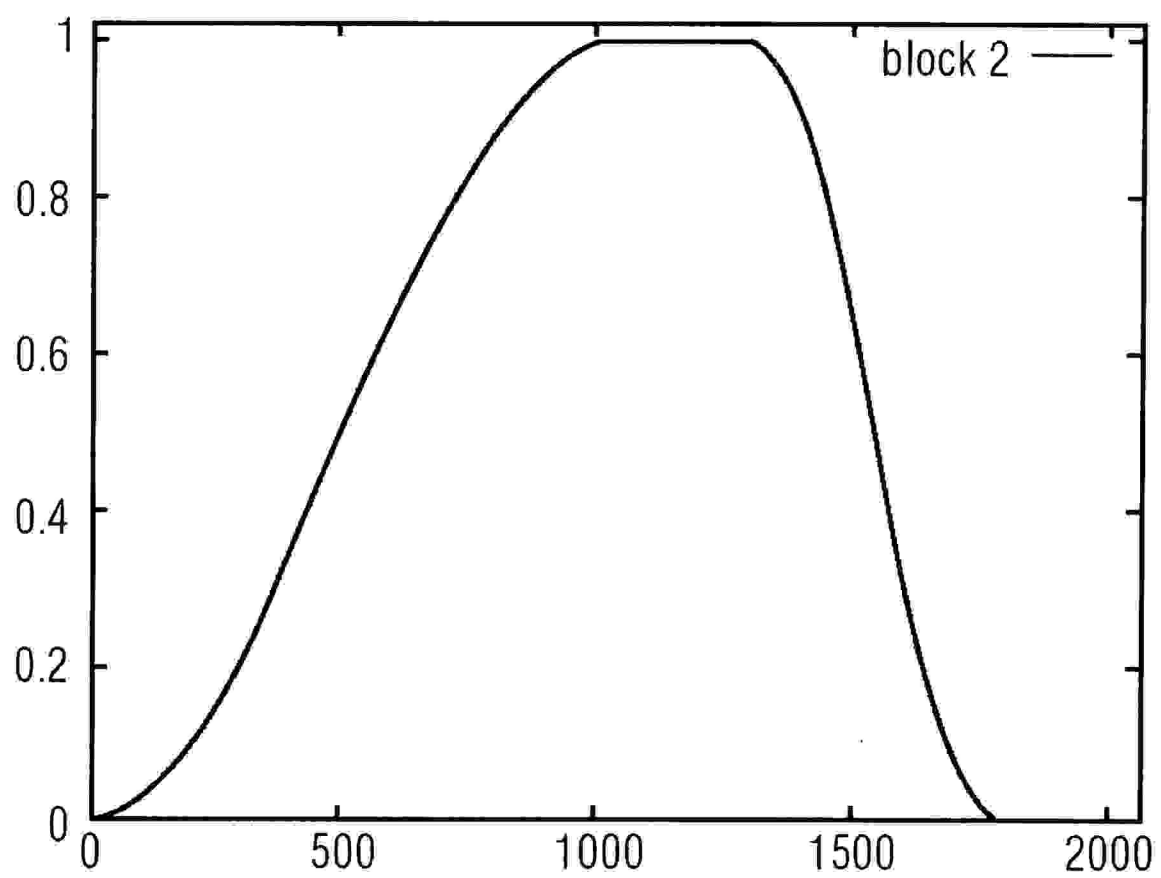


FIG 7

8/16

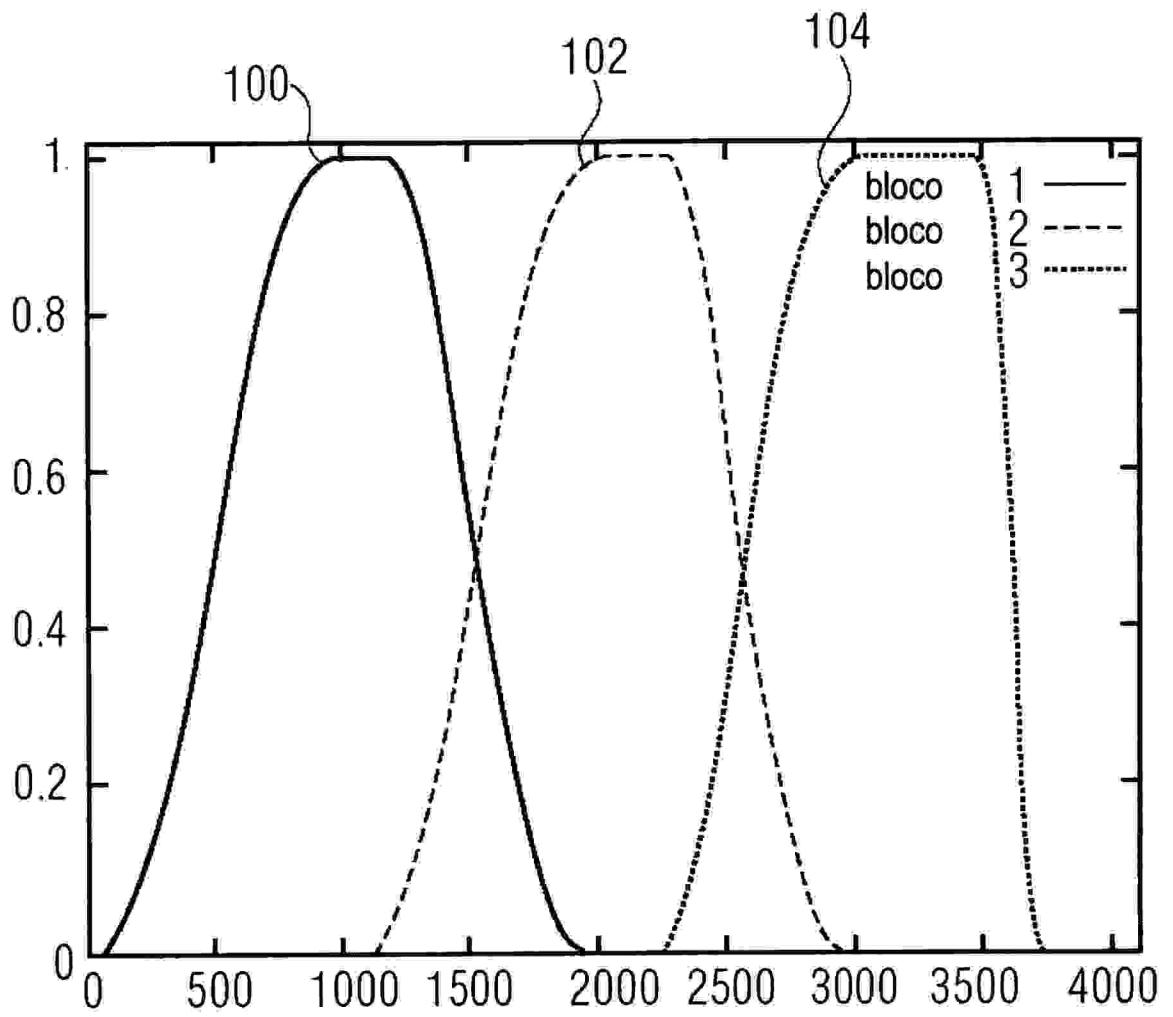


FIG 8

9/16

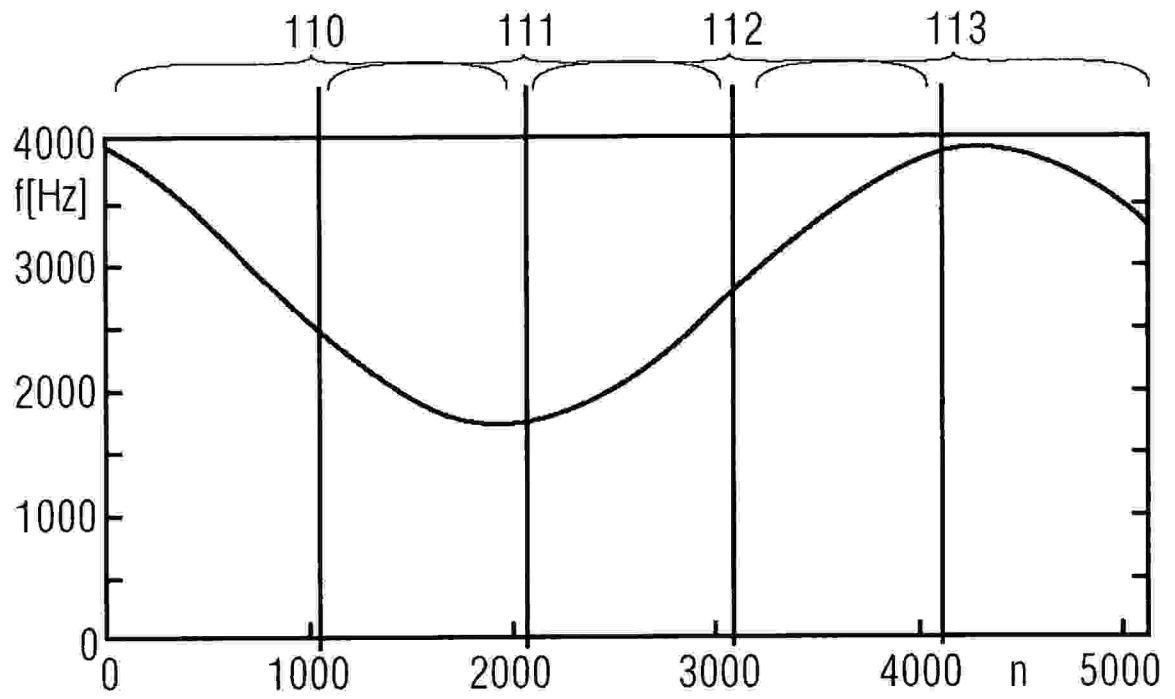


FIG 9

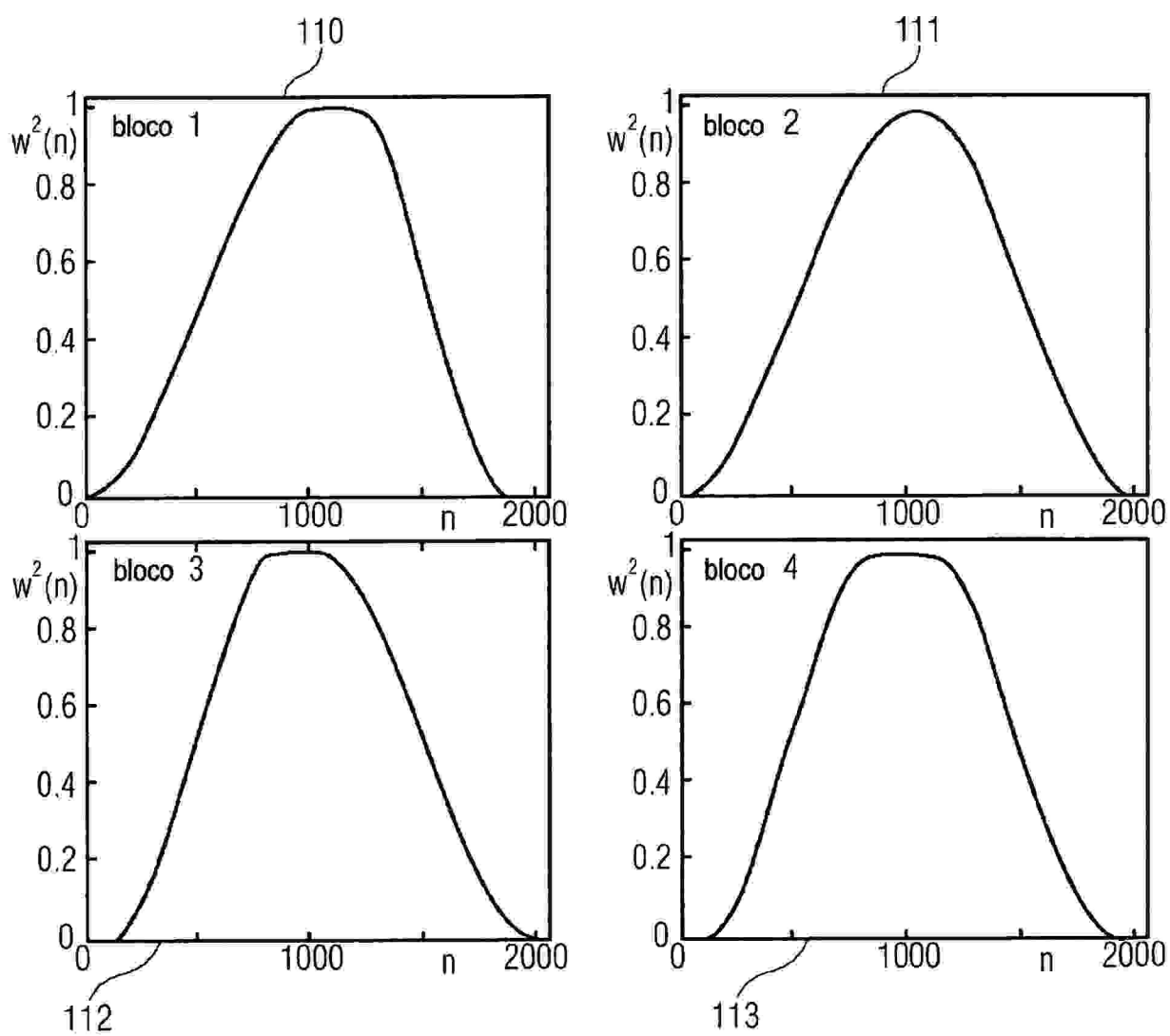


FIG 10

11/16

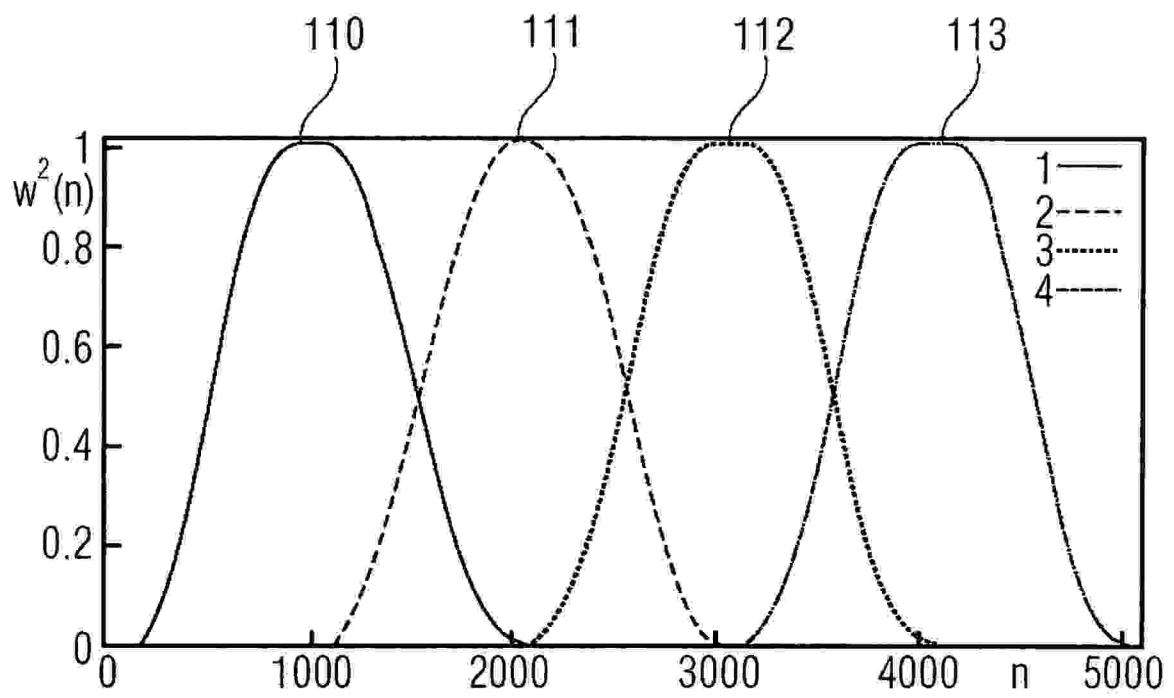


FIG 11

12/16

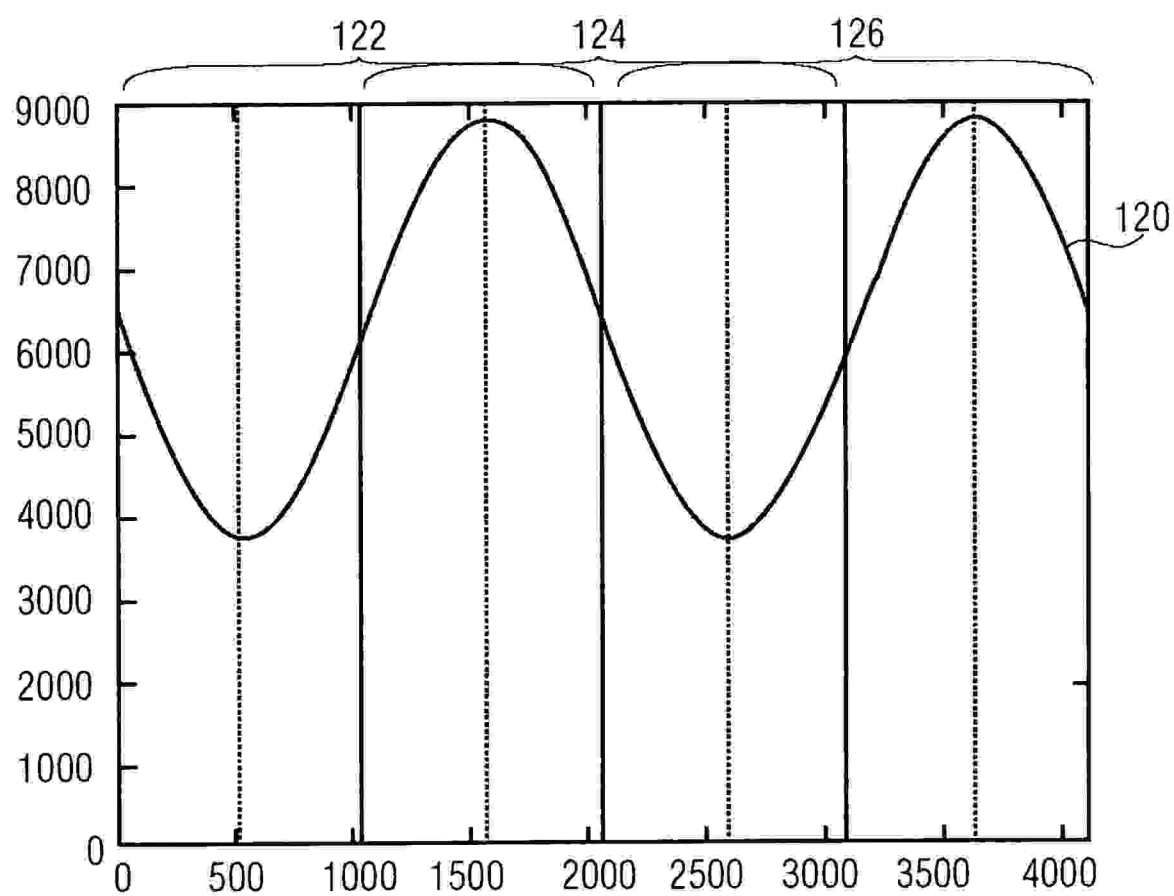


FIG 11A

13/16

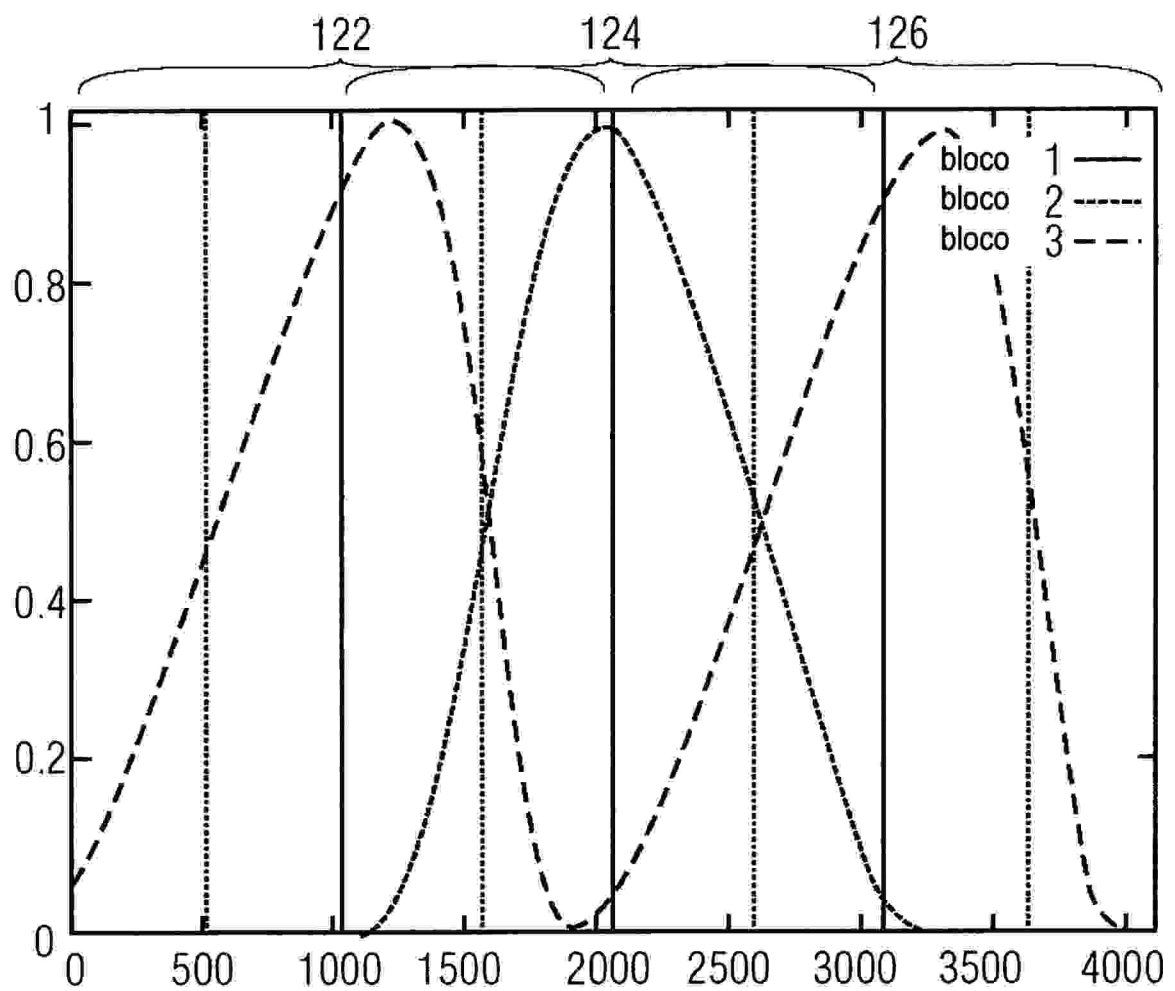


FIG 11B

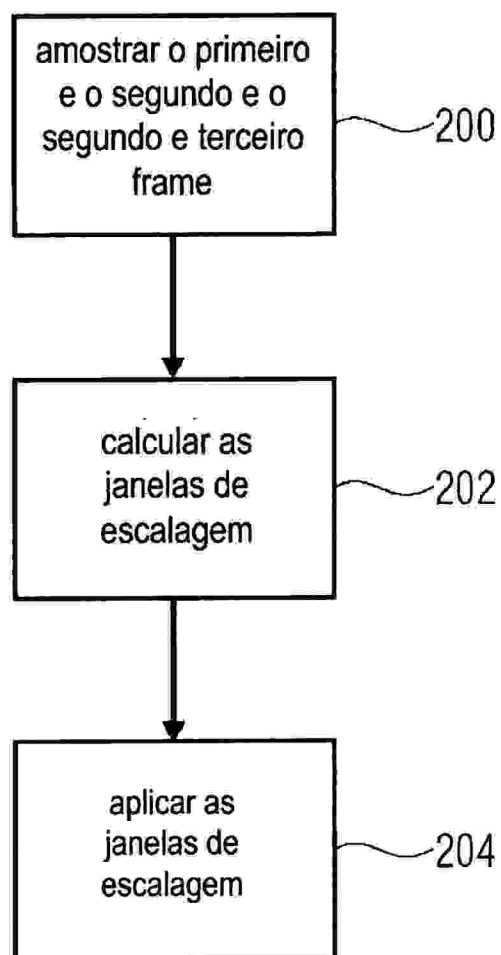
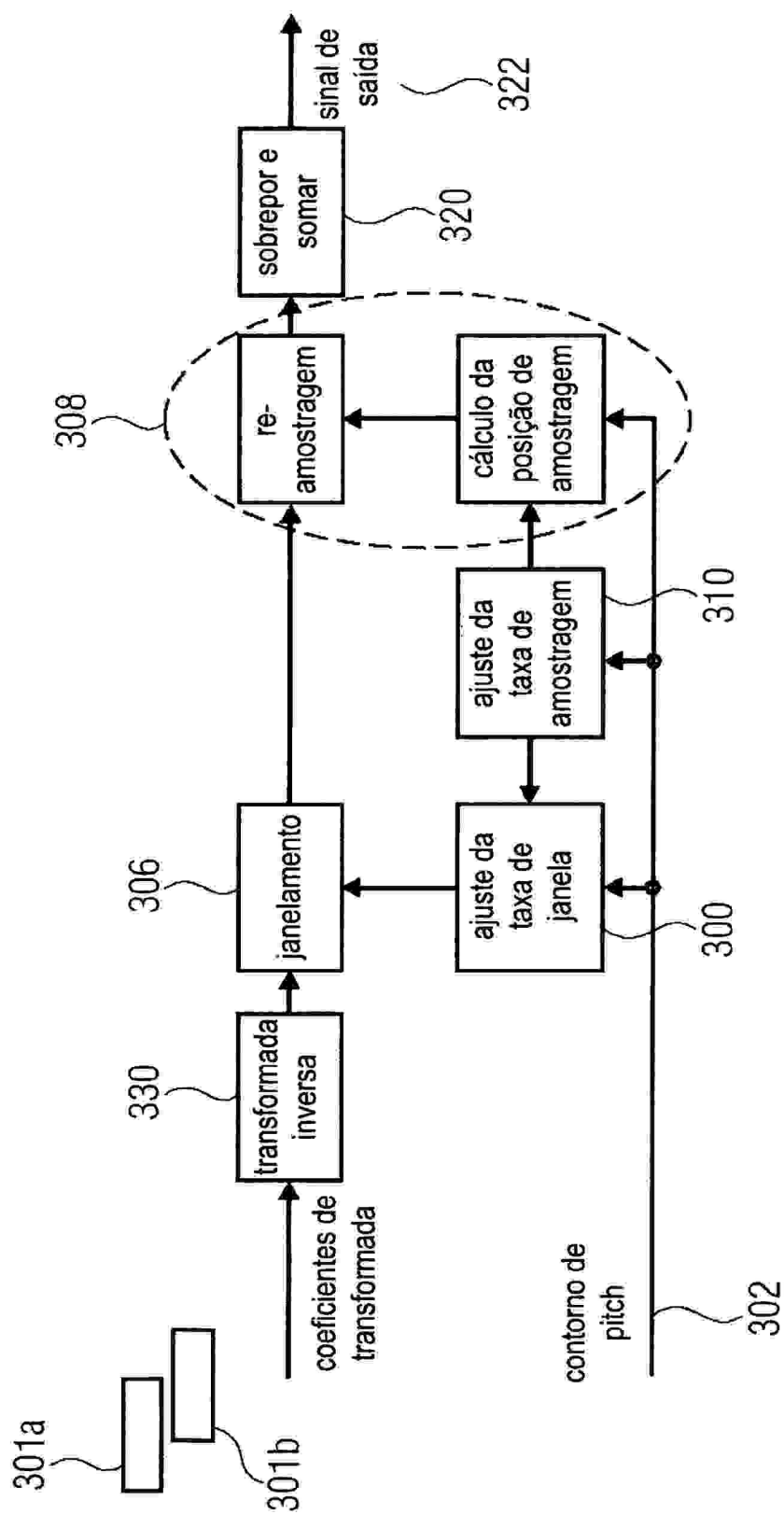


FIG 12



290

FIG 13

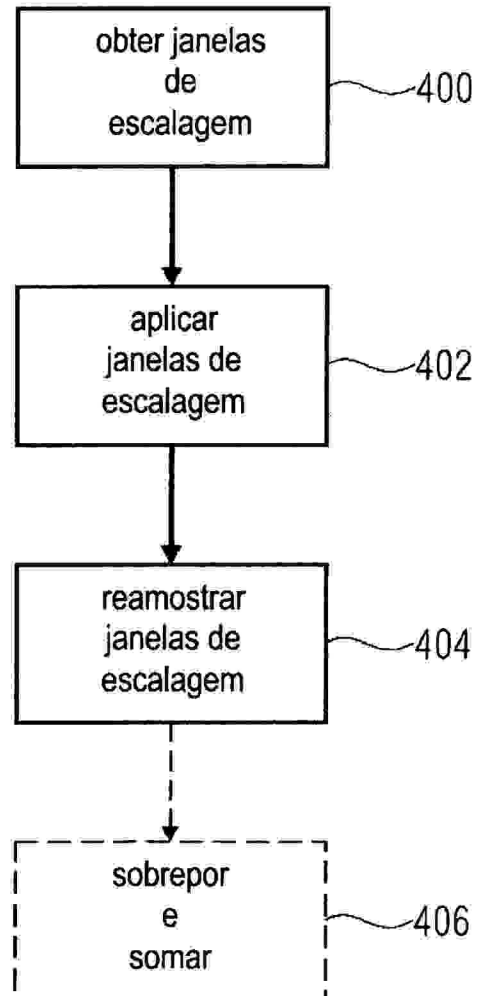


FIG 14