



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

(21) **PI0711063-4 A2**



(22) Data de Depósito: 30/03/2007
(43) Data da Publicação: 23/08/2011
(RPI 2120)

(51) *Int.Cl.:*
H03G 3/30 2006.01
H03G 7/00 2006.01

(54) Título: **CONTROLE DE GANHO DE ÁUDIO UTILIZANDO DETECÇÃO DE EVENTO AUDITIVO BASEADO EM SONORIDADE ESPECÍFICA**

(30) Prioridade Unionista: 27/04/2006 US 60/795,808

(73) Titular(es): Dolby Laboratories Licensing Corporation

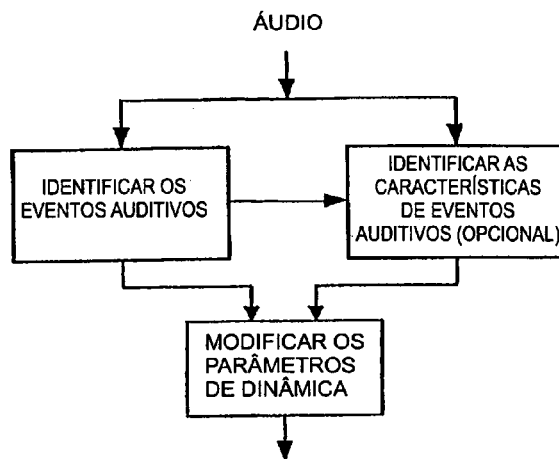
(72) Inventor(es): Alan Jeffrey Seefeldt, Brett Graham Crockett

(74) Procurador(es): Dannemann, Siemsen, Bigler & Ipanema Moreira

(86) Pedido Internacional: PCT US2007008313 de 30/03/2007

(87) Publicação Internacional: WO 2007/127023 de 08/11/2007

(57) Resumo: CONTROLE DE GANHO DE ÁUDIO UTILIZANDO DETECÇÃO DE EVENTO AUDITIVO BASEADO EM SONORIDADE ESPECÍFICA. A presente invenção refere-se a modificações de ganho dinâmico aplicadas a um sinal de áudio pelo menos parcialmente em resposta a eventos auditivos e/ou o grau de mudança em características de sinal associadas aos ditos limites de evento auditivo. Em outro aspecto, um sinal de áudio é dividido em eventos auditivos pela comparação da diferença em sonoridade específica entre os sucessivos blocos de tempo do sinal de áudio.



Relatório Descritivo da Patente de Invenção para "**CONTROLE DE GANHO DE ÁUDIO UTILIZANDO DETECÇÃO DE EVENTO AUDITIVO BASEADO EM SONORIDADE ESPECÍFICA**".

Campo da Técnica

5 A presente invenção refere-se a métodos e aparelho de controle de faixa dinâmica de áudio nos quais um dispositivo de processamento de áudio analisa um sinal de áudio e muda o nível, o ganho ou a faixa dinâmica do áudio, e todos ou alguns dos parâmetros do ganho de áudio e do processamento dinâmico são gerados como uma função de eventos auditivos. A
10 invenção também refere-se a programas de computador para praticar tais métodos ou controlar tal aparelho.

A presente invenção também refere-se a métodos e aparelho que utilizam uma detecção baseada em sonoridade específica de eventos auditivos. A invenção também refere-se a programas de computador para
15 praticar tais métodos ou controlar tal aparelho.

Antecedentes da Técnica

Processamento Dinâmico de Áudio

As técnicas de controle de ganho automático (AGC) e controle de faixa dinâmica (DRC) são bens conhecidas e são um elemento comum
20 em muitos percursos de sinal de áudio. Em um senso abstrato, ambas as técnicas medem o nível de um sinal de áudio em algum modo e então modificam em ganho o sinal por um valor que é uma função do nível medido. Em um sistema de processamento dinâmico linear, 1:1, o áudio de entrada não é processado e o sinal de áudio de saída idealmente coincide com o sinal de
25 áudio de entrada. Além disso, se tivesse um sistema de processamento dinâmico de áudio que mede automaticamente as características do sinal de entrada e utiliza esta medição para controlar o sinal de saída, se o sinal de entrada aumentar de nível por 6 dB e o sinal de saída for processado de modo que este somente aumente de nível por 3 dB, então o sinal de saída
30 foi comprimido por uma razão de 2:1 em relação ao sinal de entrada. A Publicação Internacional Número WO 2006/047600 A1 ("Calculando e Ajustando a Sonoridade Percebida e/ou o Balanço Espectral Percebido de Um Sinal

de Áudio" por Alan Jeffrey Seefeldt) provê uma vista geral detalhada dos cinco tipos básicos de processamento dinâmico de áudio: compressão, limitação, controle de ganho automático (AGC), expansão e obstrução.

Eventos Auditivos e Detecção de Evento Auditivo

5 A divisão de sons em unidades ou segmentos percebidos como separados e distintos é algumas vezes referida como "análise de evento auditivo" ou "análise de cena auditiva" ("ASA") e os segmentos são algumas vezes referido como "eventos auditivos" ou "eventos de áudio". Uma discussão extensa de análise de cena auditiva é apresentada por Albert S. Bregman no seu livro Auditory Scene Analysis - The Perceptual Organization of Sound, Massachusetts Institute of Technology, 1991, Quarta impressão, 10 2001, Segunda edição de brochura de MIT Press. Além disso, a Patente U.S. Número 6.002.776 para Bhadkamkar et al., 14 de Dezembro de 1999 cita publicações que datam até 1976 como "trabalho de técnica anterior relativo à separação de som por análise de cena auditiva". No entanto, a patente 15 de Bhadkamkar et al. desencoraja a utilização prática de análise de cena auditiva, concluindo que "[t]écnicas que envolvem a análise de cena auditiva, apesar de interessantes de um ponto de vista científico como modelos de processamento auditivo humano, são atualmente computacionalmente solici- 20 tantes demais e especializados para serem consideradas técnicas práticas para separação de som até que um progresso fundamental seja feito".

Um modo útil para identificar os eventos auditivos é apresentado por Crockett e Crocket et al. em vários pedidos de patente e documentos abaixo listados sob o cabeçalho "Incorporação Por Referência". De acordo 25 com estes documentos, um sinal de áudio está dividido em eventos auditivos, cada um dos quais tende a ser percebido como separado e distinto, pela detecção de mudanças em composição espectral (amplitude como uma função de frequência) em relação ao tempo. Isto pode ser feito, por exemplo, calculando o conteúdo espectral de blocos de tempo sucessivos do sinal de 30 áudio, calculando a diferença em conteúdo espectral entre blocos de tempo sucessivos do sinal de áudio, e identificando um limite de evento auditivo como o limite entre sucessivos blocos de tempo quando a diferença no con-

teúdo espectral entre tais blocos de tempo sucessivos excede um limite. Alternativamente, mudanças em amplitude em relação ao tempo podem ser calculadas ao invés de ou em adição a mudanças em composição espectral com relação ao tempo.

5 Em sua implementação menos computacionalmente solicitante, o processo divide o áudio em segmentos de tempo pela análise da banda de frequência inteira (áudio de largura de banda total) ou substancialmente a banda de frequência inteira (em implementações práticas, uma filtragem de limitação de banda nas extremidades do espectro é freqüentemente empre-

10 gada), e dando o maior peso para os componentes de sinal de áudio mais altos. Esta proposta aproveita-se de um fenômeno psicoacústico no qual em menores escalas de tempo (20 milissegundos (ms) e menos) o ouvido pode tender a focalizar sobre um único evento auditivo a um dado tempo. Isto implica em que apesar de múltiplos eventos poderem estar ocorrendo ao

15 mesmo tempo, um componente tende a ser perceptualmente mais proeminente e pode ser processado individualmente como se este fosse o único evento acontecendo. O aproveitamento deste efeito também permite que a detecção de evento auditivo escale com a complexidade do áudio que está sendo processado. Por exemplo, se o sinal de áudio de entrada que está

20 sendo processado for um instrumento solo, os eventos de áudio que são identificados provavelmente serão as notas individuais que estão sendo tocadas. Similarmente para um sinal de voz de entrada, os componentes individuais de fala, as vogais e as consoantes por exemplo, serão provavelmente identificados como elementos de áudio individuais. Conforme a complexi-

25 dade do áudio aumenta, tal como uma música com bateria ou múltiplos instrumentos e voz, a detecção de evento auditivo identifica o elemento de áudio "mais proeminente" (isto é, o mais alto) em qualquer dado momento.

A expensas de uma maior complexidade computacional, o processo pode também levar em consideração as mudanças em composição

30 espectral em relação ao tempo em sub-bandas de frequência discretas (sub-bandas fixas ou dinamicamente determinadas ou tanto fixas quanto dinamicamente determinadas) ao invés da largura de banda total. Esta proposta

alternativa leva em conta mais do que um fluxo de áudio em diferentes sub-bandas de frequência ao invés de assumir que somente um único fluxo é perceptível em um tempo específico.

5 Uma detecção de evento auditivo pode ser implementada dividindo uma forma de onda de áudio de domínio de tempo em intervalos ou blocos de tempo e então convertendo os dados em cada bloco para o domínio de frequência, utilizando ou um banco de filtros ou uma transformação de tempo - frequência, tal como a FFT. A amplitude do conteúdo espectral de cada bloco pode ser normalizada de modo a eliminar ou reduzir o efeito de
10 mudanças de amplitude. Cada representação de domínio de frequência resultante provê uma indicação do conteúdo espectral do áudio no bloco específico. O conteúdo espectral de blocos sucessivos é comparado e as mudanças maiores do que um limite podem ser tomadas para indicar o início temporal ou o fim temporal de um evento auditivo.

15 Preferencialmente, os dados de domínio de frequência são normalizados, como é descrito abaixo.

O grau no qual os dados de domínio de frequência precisam ser normalizados fornece uma indicação de amplitude. Com isto, se uma mudança neste grau exceder um limite predeterminado isto também pode ser
20 tomado para indicar um limite de evento. Os pontos de início e fim de evento que resultam de mudanças espectrais e de mudanças de amplitude podem ser ORed juntos de modo que os limites de evento que resultam de cada tipo de mudança sejam identificados.

25 Apesar das técnicas descritas nos ditos pedidos e documentos de Crockett e Crocket et al. serem especificamente úteis em conexão com os aspectos da presente invenção, outras técnicas para identificar os eventos auditivos e os limites de eventos podem ser empregadas em aspectos da presente invenção.

Descrição da Invenção

30 O processamento dinâmico de áudio da técnica anterior convencional envolve multiplicar o áudio por um sinal de controle que varia no tempo que ajusta o ganho do áudio produzindo um resultado desejado. "Ganho"

é um fator de escalagem que escala a amplitude de áudio. Este sinal de controle pode ser gerado em uma base contínua ou de blocos de dados de áudio, mas é geralmente derivado por alguma forma de medição do áudio que está sendo processado, e esta taxa de mudança é determinada por filtros de suavização, algumas vezes com características fixas e algumas vezes com características que variam com a dinâmica do áudio. Por exemplo, os tempos de resposta podem ser ajustáveis de acordo com as mudanças na magnitude ou na potência do áudio. Os métodos da técnica anterior tais como o controle de ganho automático (AGC) e a compressão de faixa dinâmica (DRC) não avaliam em nenhum modo psicoacusticamente baseado os intervalos de tempo durante os quais as mudanças de ganho podem ser percebidas como deteriorações e quando estas podem ser aplicadas sem imprimir artefatos audíveis. Portanto, os processos de dinâmica de áudio convencionais podem freqüentemente introduzir artefatos audíveis, isto é, os efeitos do processamento de dinâmica podem introduzir mudanças perceptíveis indesejadas no áudio.

A análise de cena auditiva identifica os eventos auditivos perceptualmente discretos, com cada evento ocorrendo entre dois limites de evento auditivo consecutivos. As deteriorações audíveis causadas por uma mudança de ganho podem ser grandemente reduzidas assegurando que dentro de um evento auditivo o ganho é mais praticamente constante e confinando muitas das mudanças à vizinhança de um limite de evento. No contexto de compressores ou expansores, a resposta a um aumento em nível de áudio (freqüentemente denominada o ataque) pode ser rápida, comparável com ou mais curta do que a duração mínima de eventos auditivos, mas a resposta a uma diminuição (a liberação ou recuperação) pode ser mais lenta do que os sons que deveriam parecer constantes ou decair gradualmente mas audivelmente perturbados. Sob tais circunstâncias, é muito benéfico retardar a recuperação de ganho até o próximo limite ou diminuir a taxa de mudança de ganho durante um evento. Para as aplicações de controle de ganho automáticas, onde o nível médio a longo prazo ou sonoridade do áudio é normalizado e tanto o tempo de ataque quanto de liberação podem portanto ser

longos comparados com a duração mínima de um evento auditivo, é benéfico durante os eventos retardar as mudanças ou diminuir as taxas de mudança em ganho até o próximo limite de evento tanto para os ganhos crescentes quanto decrescentes.

5 De acordo com um aspecto da presente invenção, um sistema de processamento de áudio recebe um sinal de áudio e analisa e altera o ganho e/ou as características de faixa dinâmica do áudio. A modificação de faixa dinâmica do áudio é freqüentemente controlada por parâmetros de um sistema de processamento de dinâmica (tempo de ataque e liberação, taxa
10 de compressão, etc.) que tem efeitos significativos sobre os artefatos perceptuais introduzidos pelo processamento de dinâmica. As mudanças em características de sinal em relação ao tempo no sinal de áudio são detectadas e identificadas como limites de evento auditivo, de modo que um segmento de áudio entre limites consecutivos constitui um evento auditivo no
15 sinal de áudio. As características dos eventos auditivos de interesse podem incluir as características dos eventos tais como força perceptual ou duração. Alguns dos ditos um ou mais parâmetros de processamento de dinâmica são gerados pelo menos parcialmente em resposta a eventos auditivos e/ou o grau de mudança em características de sinal associadas com os ditos limites
20 de evento auditivo.

Tipicamente, um evento auditivo é um segmento de áudio que tende a ser percebido como separado e distinto. Uma medida utilizável de características de sinal inclui uma medida do conteúdo espectral do áudio, por exemplo, como descrito nos documentos de Crockett e Crocket et al.
25 citados. Todos ou alguns dos um ou mais parâmetros de processamento de dinâmica de áudio podem ser gerados pelo menos parcialmente em resposta à presença ou à ausência e às características de um ou mais eventos auditivos. Um limite de evento auditivo pode ser identificado como uma mudança em características de sinal em relação ao tempo que excede um limite. Alternativamente, todos ou alguns dos um ou mais parâmetros podem ser gerados pelo menos parcialmente em resposta a uma medida continuada do
30 grau de mudança em características de sinal associadas com os ditos limites

de evento auditivo. Apesar de que, em princípio, os aspectos da invenção podem ser implementados em domínios analógicos e/ou digitais, as implementações práticas são prováveis serem implementadas no domínio digital no qual cada um dos sinais de áudio está representado por amostras individuais ou amostras dentro de blocos de dados. Neste caso, as características de sinal podem ser o conteúdo espectral de áudio dentro de um bloco, a detecção de mudanças em características de sinal em relação ao tempo podem ser a detecção de mudanças em conteúdo espectral de áudio de bloco para bloco, e os limites de início e fim temporais de evento auditivo cada um coincide com um limite de um bloco de dados. Deve ser notado que para o caso mais tradicional de executar as mudanças de ganho dinâmico em uma base de amostra por amostra, que a análise de cena auditiva descrita poderia ser executada em uma base de bloco e as informações de evento auditivo resultantes sendo utilizadas para executar as mudanças de ganho dinâmico que são aplicadas amostra por amostra.

Pelo controle dos parâmetros de processamento de dinâmica de áudio chaves utilizando os resultados de análise de cena auditiva, uma redução dramática de artefatos audíveis introduzidos por processamento de dinâmica pode ser conseguida.

A presente invenção apresenta dois modos para executar uma análise de cena auditiva. O primeiro executa um análise espectral e identifica a localização de eventos de áudio perceptíveis que são utilizados para controlar os parâmetros de ganho dinâmico pela identificação de mudanças em conteúdo espectral. O segundo modo transforma o áudio em um domínio de sonoridade perceptual (que pode prover mais informações psicoacusticamente relevantes do que o primeiro modo) e identifica a localização de eventos auditivos que são subsequentelemente utilizados para controlar os parâmetros de ganho dinâmico. Deve ser notado que o segundo modo requer que o processamento de áudio esteja ciente de níveis de reprodução acústica absolutos, o que pode não ser possível em algumas implementações. A apresentação de ambos os métodos de análise de cena auditiva permite implementações de modificação de ganho dinâmico controladas por ASA utili-

zando os processos ou dispositivos que podem ser calibrados ou não para levar em conta os níveis de reprodução absolutos.

Os aspectos da presente invenção estão aqui descritos em um ambiente de processamento de dinâmica de áudio que inclui os aspectos de
5 outras invenções. Tais outras invenções estão descritas em vários Pedidos de Patente Pendentes dos Estados Unidos ou Internacionais do Dolby Laboratories Licensing Corporation, o proprietário do presente pedido, cujos pedidos estão aqui identificados.

Descrição dos Desenhos

10 Figura 1 é um fluxograma que mostra um exemplo de etapas de processamento para executar uma análise de cena auditiva.

Figura 2 mostra um exemplo de processamento de bloco, abertura de janelas e execução de DFT em áudio enquanto executando a análise de cena auditiva.

15 Figura 3 está na natureza de um fluxograma ou diagrama de blocos funcional, que mostra um processamento paralelo no qual o áudio é utilizado para identificar os eventos auditivos e identificar as características dos eventos auditivos de modo que os eventos e suas características sejam utilizados para modificar os parâmetros de processamento de dinâmica.

20 Figura 4 está na natureza de um fluxograma ou diagrama de blocos funcional, que mostra um processamento no qual o áudio é utilizado somente para identificar os eventos auditivos e as características de evento são determinadas da detecção de evento de áudio de modo que os eventos e suas características sejam utilizados para modificar os parâmetros de pro-
25 cessamento de dinâmica.

Figura 5 está na natureza de um fluxograma ou diagrama de blocos funcional, que mostra um processamento no qual o áudio é utilizado somente para identificar os eventos auditivos e as características de evento são determinadas da detecção de evento de áudio de modo que somente as
30 características de eventos auditivos sejam utilizadas para modificar os parâmetros de processamento de dinâmica.

Figura 6 mostra um conjunto de respostas características de fil-

tro auditivo idealizadas que aproximam um requadramento crítico na escala de ERB. A escala horizontal é a frequência em Hertz e escala vertical é o nível em decibéis.

Figura 7 mostra contornos de sonoridade iguais da ISO 226. A escala horizontal é a frequência em Hertz (escala logarítmica de base 10) e a escala vertical e o nível de pressão de som em decibéis.

Figuras 8a-c mostram as características de entrada / saída idealizadas e as características de ganho de entrada de um compressor de faixa dinâmica de áudio.

Figuras 9a-f mostram um exemplo da utilização de eventos auditivos para controlar o tempo de liberação em uma implementação digital de um Controlador de Faixa Dinâmica (DRC) tradicional no qual o controle de ganho é derivado da potência de Valor Médio Quadrático (RMS) do sinal.

As figuras 10a-f mostram um exemplo da utilização de eventos auditivos para controlar o tempo de liberação em uma implementação digital de um Controlador de Faixa Dinâmica (DRC) tradicional no qual o controle de ganho é derivado da potência de Valor Médio Quadrático (RMS) do sinal para um sinal alternativo àquele utilizado na figura 9.

Figura 11 apresenta um conjunto adequado de curvas de AGC e de DRC idealizadas para a aplicação de AGC seguido por DRC em um sistema de processamento de dinâmica de domínio de sonoridade. O objetivo da combinação é fazer com que todo o áudio processado tenha aproximadamente a mesma sonoridade percebida enquanto ainda mantendo pelo menos alguma da dinâmica de áudio original.

Melhor Modo para Executar a Invenção

Análise de Cena Auditiva (Método de Domínio Original, Sem Sonoridade)

De acordo com uma modalidade de um aspecto da presente invenção, uma análise de cena auditiva pode ser composta de quatro etapas de processamento gerais como mostrado em uma porção da figura 1. A primeira etapa 1-1 ("Executar Análise Espectral") toma um sinal de áudio de domínio de tempo, divide-o em blocos e calcula um perfil espectral ou conteúdo espectral para cada um dos blocos. A análise espectral transforma o

sinal de áudio no domínio de frequência de curto prazo. Isto pode ser executado utilizando qualquer banco de filtros, ou baseado em transformadas ou bancos de filtros de passagem de banda, e em espaço de frequência ou linear ou distorcido (tal como a escala de Bark ou banda crítica, a qual melhora-se aproxima das características do ouvido humano). Com qualquer banco de

5 filtros existe uma negociação entre o tempo e a frequência. Uma maior resolução de tempo, e com isto intervalos de tempo mais curtos, leva a uma menor resolução de frequência. Uma maior resolução de frequência, e com isto sub-bandas mais estreitas, leva a intervalos de tempo mais longos.

10 A primeira etapa, ilustrada conceitualmente na figura 1, calcula o conteúdo espectral de sucessivos segmentos de tempo do sinal de áudio. Em uma modalidade prática, o tamanho de bloco de ASA pode ser que qualquer número de amostras do sinal de áudio de entrada, apesar de que 512 amostras provêm uma boa negociação de tempo e resolução de frequên-

15 cia. Na segunda etapa 1-2, as diferenças em conteúdo espectral de bloco para bloco são determinadas ("Executar medições de diferença de perfil espectral"). Assim, a segunda etapa calcula a diferença em conteúdo espectral entre os sucessivos segmentos de tempo do sinal de áudio. Como acima discutido, um indicador poderoso do início ou do fim de um evento auditivo

20 percebido acredita-se ser uma mudança em conteúdo espectral. Na terceira etapa 1-2 ("Identificar a localização de limites de evento auditivo"), quando a diferença espectral entre um bloco de perfil espectral e o seguinte é maior do que um limite, o limite de bloco é considerado ser um limite de evento auditivo. O segmento de áudio entre limites consecutivos constitui um evento au-

25 ditivo. Assim, a terceira etapa determina um limite de evento auditivo entre os sucessivos segmentos de tempo quando a diferença no conteúdo de perfil espectral entre tais segmentos de tempo sucessivos excede um limite, assim definindo os eventos auditivos. Nesta modalidade, os limites de evento auditivo definem os eventos auditivos que tem um comprimento que é um

30 múltiplo integral de blocos de perfil espectral com um comprimento mínimo de um bloco de perfil espectral (512 amostras neste exemplo). Em princípio, os limites de eventos não precisam ser tão limitados. Como uma alternativa

às modalidades práticas aqui discutidas, o tamanho de bloco de entrada pode variar, por exemplo, de modo a ter essencialmente o tamanho de um evento auditivo.

Após a identificação dos limites de eventos, as características chave do evento auditivo são identificadas, como mostrado na etapa 1-4.

Os segmentos sobrepostos ou não sobrepostos do áudio podem ser abertos em janelas e utilizados para computar os perfis espectrais do áudio de entrada. A sobreposição resulta em uma resolução mais fina quanto à localização de eventos auditivos e, também, torna menos provável perder um evento, tal como um transiente curto. No entanto, a sobreposição também aumenta a complexidade computacional. Assim, a sobreposição pode ser omitida. A figura 2 mostra uma representação conceitual de N blocos de amostra não sobrepostos sendo abertos em janelas e transformados no domínio de frequência pela Transformada de Fourier Discreta (DFT). Cada bloco pode ser aberto em janela e transformado no domínio de frequência, tal como pela utilização da DFT, de preferência implementada como uma Transformada de Fourier Rápida (FFT) para velocidade.

As seguintes variáveis podem ser utilizadas para computar o perfil espectral do bloco de entrada:

M = número de amostras abertas em janelas em um bloco utilizado para computar um perfil espectral

P = número de amostras de sobreposição de computação espectral

Em geral, quaisquer números inteiros podem ser utilizados para as variáveis acima. No entanto, a implementação será mais eficiente se M for determinado igual a uma potência de 2 de modo que FFTs padrão podem ser utilizadas para os cálculos de perfil espectral. Em uma modalidade prática do processo de análise de cena auditiva, os parâmetros listados podem ser ajustados para:

M = 512 amostras (ou 11,6 ms a 44,1 kHz)

P = 0 amostras (sem sobreposição)

Os valores acima listados foram determinados experimentalmen-

te e foram descobertos geralmente identificar com uma precisão suficiente a localização e a duração de eventos auditivos. No entanto, ajustando o valor de P para 256 amostras (50% de sobreposição) ao invés de zero amostras (sem sobreposição) foi descoberto ser útil na identificação de eventos difíceis de encontrar. Apesar de muitos diferentes tipos de janelas poderem ser utilizados para minimizar os artefatos espectrais devido à abertura em janela, a janela utilizada nos cálculos de perfil espectral é uma janela de M pontos de Hanning, Kaiser-Bessel ou outra janela adequada, de preferência não retangular. Os valores acima indicados e um tipo de janela de Hanning foram selecionados após uma extensa análise experimental já que estes mostraram prover excelentes resultados através de uma ampla faixa de material de áudio. A abertura em janela não retangular é preferida para o processamento de sinais de áudio com conteúdo de frequência predominantemente baixo. A abertura em janela retangular produz artefatos espectrais que podem causar uma detecção incorreta de eventos. Ao contrário de certas aplicações de codificador / decodificador (codec) onde um processo de sobreposição / adição total deve prover um nível constante, tal restrição não aplica-se aqui e a janela pode ser escolhida por características tais como a sua resolução de tempo / frequência e rejeição de banda de corte.

Na etapa 1-1 (figura 1), o espectro de cada bloco de amostras M pode ser computado abrindo em janela os dados com uma janela de M pontos de Hanning, Kaiser-Bessel ou outra janela adequada, convertendo para o domínio de frequência utilizando uma Transformada de Fourier Rápida de M pontos, e calculando a magnitude dos coeficientes de FFT complexos. Os dados resultantes são normalizados de modo que a maior magnitude é ajustada para a unidade, e a matriz normalizada de M números é convertida para o domínio log. Os dados podem também ser normalizados por alguma outra métrica tal como o valor de magnitude média ou o valor de potência média dos dados. A matriz não precisa ser convertida para o domínio log, mas a conversão simplifica o cálculo da medida de diferença na etapa 1-2. Mais ainda, o domínio log coincide mais proximamente com a natureza do sistema auditivo humano. Os valores de domínio log resultantes tem uma faixa de

menos infinito até zero. Uma modalidade prática, um limite inferior pode ser imposto sobre a faixa de valores; o limite pode ser fixo, por exemplo -60 dB, ou ser dependente de freqüência para refletir a audibilidade mais baixa de sons discretos em freqüências baixas e muito altas. (Note que seria possível

5 reduzir o tamanho da matriz para $M/2$ em que a FFT representa as freqüências negativas assim como as positivas).

A etapa 1-2 calcula uma medida da diferença entre os espectros de blocos adjacentes. Para cada bloco, cada um dos M (log) coeficientes espectrais da etapa 1-1 é subtraído do coeficiente correspondente para o

10 bloco precedente, e a magnitude da diferença calculada (o sinal é ignorado). Estas M diferenças são então somadas para um número. Esta medida de diferença pode também ser expressa como uma diferença média por coeficiente espectral dividindo a medida de diferença pelo número de coeficientes espectrais utilizados na soma (neste caso M coeficientes).

A etapa 1-3 identifica as localizações de limites de eventos auditivos pela aplicação de um limite à matriz de medidas de diferença da etapa 1-2 com um valor limite. Quando uma medida de diferença excede um limite, a mudança em espectro é considerada suficiente para sinalizar um novo evento e o número de bloco da mudança é gravado como um limite de evento.

15

20 Para os valores de M e P acima fornecidos e para os valores de domínio log (na etapa 1-1) expressos em unidades de dB, o limite pode ser determinado igual a 2500 se a FFT de magnitude inteira (incluindo a parte espelhada) for comparada ou 1250 se metade da FFT for comparada (como acima notado, a FFT representa as freqüências negativas assim como positivas -

25 para a magnitude da FFT, uma é a imagem de espelho da outra). Este valor foi escolhido experimentalmente e este provê uma boa detecção de limite de evento auditivo. Este valor de parâmetro pode ser mudado para reduzir (aumentar o limite) ou aumentar (diminuir o limite) a detecção de eventos.

O processo da figura 1 pode ser representado mais genericamente pelas disposições equivalentes das figuras 3, 4 e 5. Na figura 3, um

30 sinal de áudio é aplicado em paralelo a uma função de "Identificar os Eventos Auditivos" ou etapa 3-1 que divide o sinal de áudio em eventos auditivos,

cada um dos quais tende a ser percebido como separado e distinto e para uma função de "Identificar as Características de Eventos Auditivos" opcional ou etapa 3-2. O processo da figura 1 pode ser empregado para dividir o sinal de áudio em eventos auditivos e as suas características identificadas ou algum outro processo adequado pode ser empregado. As informações de evento auditivo, as quais podem ser uma identificação de limites de evento auditivo, determinada pela função ou etapa 3-1 são então utilizadas para modificar os parâmetros de processamento de dinâmica de áudio (tais como ataque, liberação, razão, etc.), conforme desejado, por uma função de "Modificar os Parâmetros de Dinâmica" ou etapa 3-3. A função de "Identificar Características" opcional ou etapa 3-3 também recebe as informações de evento auditivo. A função de "Identificar Características" ou etapa 3-3 pode caracterizar alguns ou todos os eventos auditivos por uma ou mais características. Tais características podem incluir uma identificação da sub-banda dominante do evento auditivo, como descrito em conexão com o processo da figura 1. As características podem também incluir uma ou mais características de áudio, que incluem, por exemplo, uma medida de potência do evento auditivo, uma medida de amplitude do evento auditivo, uma medida de nivelamento espectral do evento auditivo, e se o evento auditivo é substancialmente silencioso, ou outras características que ajudam a modificar os parâmetros de dinâmica de modo que os artefatos audíveis negativos do processamento sejam reduzidos ou removidos. As características podem também incluir outras características tais como se o evento auditivo inclui um transiente.

Alternativas à disposição da figura 3 estão mostradas nas figuras 4 e 5. Na figura 4, o sinal de entrada de áudio não é aplicado diretamente na função de "Identificar Características" ou etapa 4-3, mas este recebe as informações da função de "Identificar os Eventos Auditivos" ou etapa 4-1. A disposição da figura 1 é um exemplo específico de tal disposição. Na figura 5, as funções ou etapas 5-1, 5-2 e 5-3 estão dispostas em série.

Os detalhes desta modalidade prática não são críticos. Outros modos de calcular o conteúdo espectral de segmentos de tempo sucessivos

do sinal de áudio, calcular as diferenças entre os segmentos de tempo sucessivos, e determinar os limites de evento auditivo nos respectivos limites entre os sucessivos segmentos de tempo quando a diferença no conteúdo de perfil espectral entre tais segmentos de tempo sucessivos excede um limite podem ser empregados.

Análise de Cena Auditiva (Novo Método de Domínio de Sonoridade)

O Pedido Internacional sob o Tratado de Cooperação de Patentes S.N. PCT/US2005/038579, depositado em 25 de Outubro de 2005, publicado como Publicação Internacional Número WO 2006/047600 A1 ("Calculando e Ajustando a Sonoridade Percebida e/ou o Balanço Espectral Percebido de Um Sinal de Áudio" por Alan Jeffrey Seefeldt) descreve, entre outras coisas, uma medida objetiva de sonoridade percebida com base em um modelo psicoacústico. O dito pedido está por meio disto incorporado por referência em sua totalidade. Como descrito no dito pedido, de um sinal de áudio, $x[n]$, um sinal de excitação $E[b,t]$ é computado que aproxima a distribuição de energia ao longo da membrana basilar do ouvido interno na banda crítica b durante o bloco de tempo t . Esta excitação pode ser computada da Transformada de Fourier Discreta de Curto Prazo (STDFT) do sinal de áudio como segue:

$$E[b,t] = \lambda_b E[b,t-1] + (1 - \lambda_b) \sum_k |T[k]|^2 |C_b[k]|^2 |X[k,t]|^2$$

(1)

onde $X[k,t]$ representa a STDFT de $x[n]$ no bloco de tempo t e na caixa k . Note que na equação 1 t representa o tempo em unidades discretas de blocos de transformadas em oposição a uma medida contínua, tal como segundos. $T[k]$ representa a resposta de freqüência de um filtro que simula a transmissão de áudio através do ouvido externo e médio, e $C_b[k]$ representa a resposta de freqüência da membrana basilar em uma localização que corresponde à banda crítica b . A figura 6 apresenta um conjunto adequado de respostas de filtro de banda crítica no qual 40 bandas estão espaçadas uniformemente ao longo da escala de Largura de Banda Retangular Equivalen-

te (ERB), como definido por Moore e Glasberg. Cada forma de filtro está descrita por uma função exponencial arredondada e as bandas estão distribuídas utilizando um espaçamento de 1 ERB. Finalmente, a constante de tempo de suavização λ_b na equação 1 pode ser vantajosamente escolhida

5 proporcional ao tempo de integração de percepção de sonoridade humana dentro da banda b .

Utilizando contornos de sonoridade iguais, tais como aqueles apresentados na figura 7, a excitação em cada banda é transformada em um nível de excitação que geraria a mesma sonoridade percebida a 1 kHz. A sonoridade específica, uma medida de sonoridade perceptual distribuída

10 através de frequência e tempo, é então computada da excitação transformada, $E_{1kHz}[b,t]$, através de uma não linearidade compressiva. Uma tal função adequada para computar a sonoridade específica $N[b,t]$ é dada por:

$$N[b,t] = \beta \left(\left(\frac{E_{1kHz}[b,t]}{TQ_{1kHz}} \right)^\alpha - 1 \right)$$

(2)

onde TQ_{1kHz} é o limite em silêncio a 1 kHz e as constantes β e α são escolhidas para coincidir com o crescimento de dados de sonoridade como coletados de experiências de escuta. Abstratamente, esta transformação de excitação para sonoridade específica pode ser representada pela função $\Psi\{ \}$ de modo que:

15

$$N[b,t] = \Psi\{E[b,t]\}$$

Finalmente, a sonoridade total, $L[t]$, representada em unidades

20 de sone, é computada somando a sonoridade específica através das bandas:

$$L[t] = \sum_b N[b,t]$$

(3)

A sonoridade específica $N[b,t]$ é uma representação espectral destinada a simular o modo no qual um humano percebe o áudio como uma função de frequência e tempo. Esta captura as variações em sensibilidade a diferentes frequências, variações em sensibilidade a nível, e variações em resolução de frequência. Como tal, esta é uma representação espectral bem coincidente com a detecção de eventos auditivos. Apesar de mais computacionalmente complexo, comparar a diferença de $N[b,t]$ através das bandas entre os sucessivos blocos de tempo pode em muitos casos resultar em uma detecção perceptualmente mais precisa de eventos auditivos em comparação com a utilização direta de sucessivos espectros de FFT acima descrito.

No dito pedido de patente, diversas aplicações para modificar o áudio com base neste modelo de sonoridade psicoacústico estão descritas. Entre estas estão diversos algoritmos de processamento de dinâmica, tal como AGC e DRC. Estes algoritmos descritos podem beneficiar-se da utilização de eventos auditivos para controlar os vários parâmetros associados. Como a sonoridade específica já está computada, esta está prontamente disponível para o propósito de detectar os ditos eventos. Os detalhes de uma modalidade preferida estão abaixo discutidos.

Controle de Parâmetro de Processamento de Dinâmica de Áudio com Eventos Auditivos

Dois exemplos de modalidades da invenção serão agora apresentados. O primeiro descreve a utilização de eventos auditivos para controlar o tempo de liberação em uma implementação digital de um Controlador de Faixa Dinâmica (DRC) no qual o controle de ganho é derivado da potência de Valor Médio Quadrático (RMS) do sinal. A segunda modalidade descreve a utilização de eventos auditivos para controlar certos aspectos de uma combinação mais sofisticada de AGC e DRC implementada dentro do contexto do modelo de sonoridade psicoacústico acima descrito. Estas duas modalidades destinam-se a servir como exemplos da invenção somente, e não deve ser compreendido que a utilização de eventos auditivos para controlar os parâmetros de um algoritmo de processamento de dinâmica não está restrita às especificidades abaixo descritas.

Controle de Faixa Dinâmica

A implementação digital descrita de um DRC segmenta um sinal de áudio $x[n]$ em blocos abertos em janela, semi-sobrepostos, e para cada bloco um ganho de modificação com base em uma medida da potência local do sinal e uma curva de compressão selecionada é computado. O ganho é suavizado através dos blocos e então multiplicado com cada bloco. Os blocos modificados são finalmente adicionados sobrepostos para gerar o sinal de áudio modificado $y[n]$.

Deve ser notado, que apesar da análise de cena auditiva e da implementação digital de DRC como aqui descrito dividir o sinal de áudio de domínio de tempo em blocos para executar a análise e o processamento, o processamento de DRC não precisa ser executado utilizando a segmentação de blocos. Por exemplo, a análise de cena auditiva poderia ser executada utilizando a segmentação de blocos e análise espectral como acima descrito e as localizações e características de eventos auditivos resultantes poderiam ser utilizadas para prover as informações de controle para uma implementação digital de uma implementação de DRC tradicional que tipicamente opera em uma base de amostra por amostra. Aqui, no entanto, a mesma estrutura de bloqueio utilizada para a análise de cena auditiva é empregada para o DRC para simplificar a descrição de sua combinação.

Prosseguindo com a descrição de uma implementação de DRC baseada em blocos, os blocos sobrepostos do sinal de áudio podem ser representados como:

$$x[n, t] = w[n]x[n + tM / 2]$$

(4)

para $0 < n < M - 1$

onde M é o comprimento de bloco e o tamanho de salto é $M/2$, $w[n]$ é a janela, n é o índice de amostra dentro do bloco, e t é o índice de bloco (note que aqui t é utilizado no mesmo modo como com a STDFT na equação 1; este representa o tempo em unidades discretas de blocos ao invés de segundos, por exemplo). Idealmente, a janela $w[n]$ afila para zero em ambas as extremidades e soma à unidade quando semi-sobreposta com si mesma; a janela

de seno comumente utilizada atende estes critérios, por exemplo.

Para cada bloco, pode-se então computar a potência de RMS para gerar uma medida de potência $P[t]$ em dB por bloco:

$$P[t] = 10 * \log_{10} \left(\frac{1}{M} \sum_{n=1}^M x^2[n, t] \right)$$

(5)

Como anteriormente mencionado, poderia suavizar-se esta medida de potência com um rápido ataque e uma lenta liberação antes do processamento com uma curva de compressão, mas como uma alternativa a potência instantânea $P[t]$ é processada e o ganho resultante é suavizado. Esta proposta alternativa tem a vantagem que uma simples curva de compressão com pontos de cotovelo agudos pode ser utilizada, mas os ganhos resultantes são ainda suavizados conforme a potência desloca-se através do ponto de cotovelo. Representando uma curva de compressão como mostrado na figura 8c como uma função F de nível de sinal que gera um ganho, o ganho de bloco $G[t]$ é dado por:

$$G[t] = F\{P[t]\}$$

(6)

Assumindo que a curva de compressão aplica uma maior atenuação conforme o nível de sinal aumenta, o ganho será decrescente quando o sinal estiver em "modo de ataque" e crescente quando no "modo de liberação". Portanto, um ganho suavizado $\bar{G}[t]$ pode ser computado de acordo com:

$$\bar{G}[t] = \alpha[t] \cdot \bar{G}[t-1] + (1 - \alpha[t]) G[t]$$

(7a)

onde

$$\alpha[t] = \begin{cases} \alpha_{attach} & G[t] < \bar{G}[t-1] \\ \alpha_{release} & G[t] \geq \bar{G}[t-1] \end{cases} \quad (7b)$$

e

$$\alpha_{release} \gg \alpha_{attach} \quad (7c)$$

Finalmente, o ganho suavizado $\bar{G}[t]$, o qual é em dB, é aplicado a cada bloco do sinal, e os blocos modificados são adicionados sobrepostos para produzir o áudio modificado:

$$y[n + tM/2] = (10^{\bar{G}[t]/20})x[n, t] + (10^{\bar{G}[t-1]/20})x[n + M/2, t-1] \quad \text{for } 0 < n < M/2 \quad (8)$$

5 Note que como os blocos foram multiplicados por uma janela afinada, como mostrado na equação 4, a síntese de adição de sobreposição acima mostrada efetivamente suaviza os ganhos através de amostras do sinal processado $y[n]$. Assim, o sinal de controle de ganho recebe suavização além daquela mostrada na equação 7a. Em uma implementação mais tradicional de DRC operando amostra por amostra ao invés de bloco por bloco, uma suavização de ganho mais sofisticada do que o simples filtro de um pólo mostrado na equação 7a poderia ser necessária de modo a impedir uma distorção audível no sinal processado. Também, a utilização de um processamento baseado em bloco introduz um retardo inerente de $M/2$ amostras no sistema, e desde que o tempo de decaimento associado com α_{attach} fique próximo deste retardo, o sinal $x[n]$ não precisa ser retardado adicionalmente antes da aplicação dos ganhos para os propósitos de impedir exceder.

20 As figuras 9a até 9c apresentam o resultado de aplicar o processamento de DRC descrito a um sinal de áudio. Para esta implementação específica, um comprimento de bloco de $M = 512$ é utilizado na taxa de amostragem de 44,1 kHz. Uma curva de compressão similar àquela mostrada

na figura 8b é utilizada:

acima de -20 dB em relação a digital em escala total o sinal é atenuado com uma razão de 5:1, e

abaixo de -30 dB o sinal é intensificado com uma razão de 5:1.

5 O ganho é suavizado com um coeficiente de ataque α_{attack} que corresponde a um meio tempo de decaimento de 10 ms e um coeficiente de liberação α_{release} que corresponde a um meio tempo de decaimento de 500 ms. O sinal de áudio original apresentado na figura 9a consiste em seis acordes de piano consecutivos, com o acorde final localizado ao redor da amostra de $1,75 \times 10^5$, decaindo para o silêncio. Examinando um gráfico do ganho $G-[t]$ na figura 9b, deve ser notado que o ganho permanece próximo de 0 dB enquanto os seis acordes são tocados. Isto é porque a energia de sinal permanece, pela maior parte, entre -30 dB e -20 dB, a região dentro da qual a curva de DRC não pede nenhuma modificação. No entanto, após a execução do último acorde, a energia de sinal cai abaixo de -30 dB, e o ganho começa a aumentar, eventualmente além de 15 dB, conforme o acorde decai. A figura 9c apresenta o sinal de áudio modificado resultante, e pode-se ver que a cauda do acorde final é intensificada significativamente. Audivelmente, esta intensificação do som de decaimento de baixo nível, natural do acorde cria um resultado extremamente não natural. É o objetivo da presente invenção impedir os problemas deste tipo que estão associados com um processador de dinâmica tradicional.

As figuras 10a até 10c apresentam os resultados da aplicação do mesmo sistema de DRC exato a um sinal de áudio diferente. Neste caso a primeira metade do sinal consiste uma peça de música de tempo crescente em um alto nível, e então a aproximadamente a amostra 10×10^4 o sinal muda para uma segunda peça de música de tempo crescente, mas a um nível significativamente mais baixo. Examinando o ganho na figura 6b, vê-se que o sinal é atenuado por aproximadamente 10 dB durante a primeira metade, e então o ganho sobe novamente para 0 dB durante a segunda metade quando a peça mais suave está tocando. Neste caso, o ganho comporta-se como desejado. Se desejaria que a segunda peça fosse intensificada em

relação à primeira, e o ganho deveria aumentar rapidamente após a transição para a segunda peça para ser audivelmente não obstrusiva. Vê-se um comportamento de ganho que é similar àquele para o primeiro sinal discutido, mas aqui o comportamento é desejável. Portanto, se gostaria de fixar o primeiro caso sem afetar o segundo. A utilização de eventos auditivos para controlar o tempo de liberação deste sistema de DRC provê tal solução.

No primeiro sinal que foi examinado na figura 9, a intensificação do decaimento do último acorde parece não natural porque o acorde e o seu decaimento são percebidos com um único evento auditivo cuja integridade é esperada ser mantida. No segundo caso, no entanto, muitos eventos auditivos ocorrem enquanto o ganho aumenta, significando que para qualquer evento individual, pouca mudança é imposta. Portanto a mudança de ganho total não é tão discutível. Pode-se portanto questionar que uma mudança de ganho deveria ser permitida somente na vizinhança quase temporal de um limite de evento auditivo. Pode-se aplicar este princípio a um ganho enquanto este está ou no modo de ataque ou de liberação, mas para a maioria das implementações práticas de um DRC, o ganho move-se tão rapidamente no modo de ataque em comparação com a resolução temporal humana de percepção de eventos que nenhum controle é necessário. Pode-se portanto utilizar os eventos para controlar a suavização do ganho de DRC somente quando este está no modo de liberação.

Um comportamento adequado do controle de liberação será agora descrito. Em termos qualitativos, se um evento for detectado, o ganho é suavizado com o tempo de liberação constante como acima especificado na Equação 7a. Conforme o tempo evolui passando pelo evento detectado, e se nenhum evento subsequente for detectado, a constante de tempo de liberação aumenta continuamente de modo que eventualmente o ganho suavizado é "congelado" no lugar. Se outro evento for detectado, então a constante de tempo de suavização é reajustada para o valor original e o processo repete-se. De modo a modular o tempo de liberação, pode-se primeiro gerar um sinal de controle baseado nos limites de evento detectados.

Como anteriormente discutido, os limites de evento podem ser

detectados procurando por mudanças em espectros sucessivos do sinal de áudio. Nesta implementação específica, a DFT de cada bloco sobreposto $x[n,t]$ pode ser computada para gerar a STDFT do sinal de áudio $x[n]$:

$$X[k,t] = \sum_{n=0}^{M-1} x[n,t] e^{-j \frac{2\pi kn}{M}} \quad (9)$$

A seguir, a diferença entre os espectros de magnitude log normalizados de blocos sucessivos pode ser computada de acordo com:

$$D[t] = \sum_k |X_{NORM}[k,t] - X_{NORM}[k,t-1]| \quad (10a)$$

onde

$$X_{NORM}[k,t] = \log \left(\frac{|X[k,t]|}{\max_k \{|X[k,t]|\}} \right) \quad (10b)$$

Aqui o máximo de $|X[k,t]|$ através das caixas k é utilizado para a normalização, apesar de que se poderia empregar outros fatores de normalização; por exemplo, a média de $|X[k,t]|$ através das caixas. Se a diferença $D[t]$ exceder um limite D_{min} , então um evento é considerado ter ocorrido. Além disso, pode-se atribuir uma força a este evento, que fica entre zero e um, com base no tamanho de $D[t]$ em comparação com um limite máximo D_{max} . O sinal de força de evento auditivo resultante $A[t]$ pode ser computado como:

$$A[t] = \begin{cases} 0 & D[t] \leq D_{min} \\ \frac{D[t] - D_{min}}{D_{max} - D_{min}} & D_{min} < D[t] < D_{max} \\ 1 & D[t] \geq D_{max} \end{cases} \quad (11)$$

Atribuindo uma força para o evento auditivo proporcional à quantidade de mudança espectral associada com aquele evento, um maior controle sobre o processamento de dinâmica é conseguido em comparação com uma decisão de evento binário. Os inventores descobriram que maiores mudanças de ganho são aceitáveis durante os eventos mais fortes, e o sinal na equação 11 permite tal controle variável.

O sinal $A[t]$ é um sinal impulsivo com um impulso ocorrendo na localização de um limite de evento. Para os propósitos de controlar o tempo de liberação, pode-se adicionalmente suavizar o sinal $A[t]$ de modo que este decaia suavemente para zero após a detecção de um limite de evento. O sinal de controle de evento suavizado $\bar{A}[t]$ pode ser computado de $A[t]$ de acordo com:

$$\bar{A}[t] = \begin{cases} A[t] & A[t] > \alpha_{event} \bar{A}[t-1] \\ \alpha_{event} \bar{A}[t-1] & \text{de outra forma} \end{cases} \quad (12)$$

Aqui α_{event} controla o tempo de decaimento do sinal de controle de evento. As figuras 9d e 10d apresentam o sinal de controle de evento $\bar{A}[t]$ para os dois sinais de áudio correspondentes, com o tempo de meio decaimento do mais suave ajustado para 250 ms. No primeiro caso, vê-se que um limite de evento é detectado para cada um dos seis acordes de piano, e que o sinal de controle de evento decai suavemente na direção de zero após cada evento. Para o segundo sinal, muitos eventos são detectados muito próximos uns dos outros no tempo, e portanto o sinal de controle de evento nunca decai totalmente para zero.

Pode-se agora utilizar o sinal de controle de evento $\bar{A}[t]$ para variar a constante de tempo de liberação utilizada para suavizar o ganho. Quando o sinal de controle é igual a um, o coeficiente de suavização $\alpha[t]$ da Equação 7a é igual a $\alpha_{release}$, como antes, e quando o sinal de controle é igual a zero, o coeficiente é igual a um de modo que o ganho suavizado é impedido de mudar. O coeficiente de suavização é interpolado entre estes

dois extremos utilizando o sinal de controle de acordo com:

$$\alpha[t] = \begin{cases} \alpha_{attack} & G[t] < \bar{G}[t-1] \\ \bar{A}[t]\alpha_{release} + (1 - \bar{A}[t]) & G[t] \geq \bar{G}[t-1] \end{cases} \quad (13)$$

Pela interpolação do coeficiente de suavização continuamente como uma função do sinal de controle de evento, o tempo de liberação é reajustado para um valor proporcional à força de evento no início de um evento e então aumenta suavemente para infinito após a ocorrência de um evento. A taxa deste aumento é ditada pelo coeficiente α_{event} utilizado para gerar o sinal de controle de evento suavizado.

As figuras 9e e 10e mostram o efeito de suavização do ganho com o coeficiente controlado por evento da Equação 13 em oposição ao coeficiente não controlado por evento da Equação 7b. No primeiro caso, o sinal de controle de evento cai para zero após o último acorde de piano por meio disto impedindo que o ganho mova-se para cima. Como um resultado, o áudio modificado correspondente na figura 9f não sofre uma intensificação não natural do decaimento do acorde. No segundo caso, o sinal de controle de evento nunca se aproxima de zero, e portanto o sinal de ganho suavizado é muito pouco inibido através da aplicação do controle de evento. A trajetória do ganho suavizado é quase idêntica ao ganho não controlado por evento na figura 10b. Este é exatamente o efeito desejado.

AGC e DRC baseados em sonoridade

Como uma alternativa às técnicas de processamento de dinâmica tradicionais onde as modificações de sinal são uma função direta de simples medições de sinal tal como o Pico ou a potência de RMS, o Pedido de Patente Internacional S.N. PCT/US2005/038579 descreve a utilização do modelo de sonoridade baseado em psicoacústica anteriormente descrito como uma estrutura dentro da qual executar o processamento de dinâmica. Diversas vantagens são citadas. Primeiro, as medições e as modificações são especificadas em unidades de sone, a qual é uma medida mais precisa de percepção de sonoridade do que as medições mais básicas tais como o

Pico ou a potência de RMS. Segundamente, o áudio pode ser modificado de modo que o balanço espectral percebido do áudio original é mantido conforme a sonoridade total é mudada. Deste modo, mudanças na sonoridade total tornam-se menos perceptualmente aparentes em comparação com um processador de dinâmica que utiliza um ganho de banda larga, por exemplo, para modificar o áudio. Finalmente, o modelo psicoacústico é inerentemente de múltiplas bandas, e portanto o sistema é facilmente configurado para executar um processamento de dinâmica de múltiplas bandas de modo a aliviar os problemas de bombeamento espectral cruzado bem conhecido associados com um processador de dinâmica de banda larga.

Apesar da execução de processamento de dinâmica neste domínio de sonoridade já apresenta diversas vantagens sobre um processamento de dinâmica mais tradicional, a técnica pode ser adicionalmente aperfeiçoada através da utilização de eventos auditivos para controlar vários parâmetros. Considere o segmento de áudio que contém os acordes de piano como apresentado em 27a e o DRC associado mostrado nas figuras 10b e c. Poderia-se executar um DRC similar no domínio de sonoridade, e neste caso, quando a sonoridade do decaimento do acorde do piano final é intensificada, a intensificação seria menos aparente porque o balanço espectral da nota em decaimento seria mantido conforme a intensificação é aplicada. No entanto, uma melhor solução é não intensificar o decaimento, e portanto pode-se vantajosamente aplicar o mesmo princípio de controlar os tempos de ataque e de liberação com os eventos auditivos no domínio de sonoridade como foi anteriormente descrito para o DRC tradicional.

O sistema de processamento de dinâmica de domínio de sonoridade que é agora descrito consiste em AGC seguido por DRC. O objetivo desta combinação é fazer com que todo o áudio processado tenha aproximadamente a mesma sonoridade percebida enquanto ainda mantendo pelo menos alguma da dinâmica de áudio original. A figura 11 apresenta um conjunto adequado de curvas de AGC e de DRC para esta aplicação. Note que a entrada e a saída de ambas as curvas estão representadas em unidades de sone já que o processamento é executado no domínio de sonoridade. A

curva de AGC esforça-se para trazer o áudio de saída mais próximo a algum nível alvo, e, como anteriormente mencionado, o faz com constantes de tempo relativamente lentas. Pode-se pensar do AGC como fazendo a sonoridade de longo prazo do áudio igual ao alvo, mas em uma base de curto prazo, a sonoridade pode flutuar significativamente ao redor deste alvo. Portanto, pode-se empregar um DRC que atue mais rápido para limitar estas flutuações a alguma faixa considerada aceitável para a aplicação específica. A figura 11 mostra tal curva de DRC onde o alvo de AGC cai dentro da "banda nula" do DRC, a porção da curva que não requer nenhuma modificação. Com esta combinação de curvas, o AGC coloca a sonoridade de longo prazo do áudio dentro da banda nula da curva de DRC de modo que modificações de DRC de atuação rápida mínimas precisam ser aplicadas. Se a sonoridade de curto prazo ainda flutuar fora da banda nula, o DRC então atua para mover a sonoridade do áudio na direção de sua banda nula. Como uma nota geral final, pode-se aplicar o AGC de atuação lenta de modo que todas as bandas do modelo de sonoridade recebam a mesma quantidade de modificação de sonoridade, por meio disto mantendo o balanço espectral percebido, e pode-se aplicar o DRC de atuação rápida em um modo que permita que a modificação de sonoridade varie através das bandas de modo a aliviar o bombeamento espectral cruzado que poderia de outro modo resultar da modificação de sonoridade independente de banda de atuação rápida.

Os eventos auditivos podem ser utilizados para controlar o ataque e a liberação tanto do AGC quanto do DRC. No caso do AGC, tanto o tempo de ataque quanto de liberação são grandes em comparação com a resolução temporal de percepção de evento, e portanto o controle de evento pode ser vantajosamente empregado em ambos os casos. Com o DRC, o ataque é relativamente curto, e portanto o controle de evento pode ser necessário somente para a liberação como com o DRC tradicional acima descrito.

Como anteriormente descrito, pode-se utilizar o espectro de sonoridade específico associado com o modelo de sonoridade empregado para os propósitos de detecção de evento. Um sinal de diferença $D[t]$, similar à-

quele nas Equações 10a e b pode ser computado da sonoridade específica $N[b,t]$, definida na Equação 2, como segue:

$$D[t] = \sum_b |N_{NORM}[b,t] - N_{NORM}[b,t-1]|$$

(14a)

onde

$$N_{NORM}[b,t] = \frac{N[b,t]}{\max_b \{N[b,t]\}}$$

(14b)

Aqui o máximo de $|N[b,t]|$ através das bandas de frequência b é utilizado para a normalização, apesar de que se poderia empregar outros fatores de normalização; por exemplo, a média de $|N[b,t]|$ através das bandas de frequência. Se a diferença $D[t]$ exceder um limite D_{min} , então um evento é considerado ter ocorrido. O sinal de diferença pode então ser processado no mesmo modo mostrado nas Equações 11 e 12 para gerar um sinal de controle de evento suave $\bar{A}[t]$ utilizado para controlar os tempos de ataque e de liberação.

A curva de AGC apresentada na figura 11 pode ser representada como uma função que toma como a sua entrada uma medida de sonoridade e gera uma sonoridade de saída desejada:

$$L_o = F_{AGC} \{L_i\}$$

(15a)

A curva de DRC pode ser similarmente representada:

$$L_o = F_{DRC} \{L_i\}$$

(15b)

Para o AGC, a sonoridade de entrada é uma medida da sonoridade de longo prazo do áudio. Pode-se computar tal medida suavizando a

sonoridade instantânea $L[t]$, definida na Equação 3, utilizando constantes de tempo relativamente longas (na ordem de diversos segundos). Foi mostrado que no julgamento de sonoridade de longo prazo de um segmento de áudio, os seres humanos ponderam as porções mais altas mais pesadamente do que as suaves, e pode-se utilizar um ataque mais rápido do que a liberação na suavização para simular este efeito. Com a incorporação de controle de evento tanto para o ataque quanto a liberação, a sonoridade de longo prazo utilizada para determinar a modificação de AGC pode portanto ser computada de acordo:

$$L_{AGC}[t] = \alpha_{AGC}[t]L_{AGC}[t-1] + (1 - \alpha_{AGC}[t])L[t] \quad (16a)$$

10 onde

$$\alpha_{AGC}[t] = \begin{cases} \bar{A}[t]\alpha_{AGCattack} + (1 - \bar{A}[t]) & L[t] > L_{AGC}[t-1] \\ \bar{A}[t]\alpha_{AGCrelease} + (1 - \bar{A}[t]) & L[t] \leq L_{AGC}[t-1] \end{cases} \quad (16b)$$

Além disso, pode ser computar um espectro de sonoridade específica de longo prazo associado que posteriormente será utilizado para o DRC de múltiplas bandas:

$$N_{AGC}[b,t] = \alpha_{AGC}[t]N_{AGC}[b,t-1] + (1 - \alpha_{AGC}[t])N[b,t] \quad (16c)$$

Na prática pode-se escolher os coeficientes de suavização de modo que o tempo de ataque seja aproximadamente a metade daquele da liberação. Dada a medida de sonoridade de longo prazo, pode-se então computar a escalagem de modificação de sonoridade associada com o AGC como a razão da sonoridade de saída para a sonoridade de entrada:

$$S_{AGC}[t] = \frac{F_{AGC} \{L_{AGC}[t]\}}{L_{AGC}[t]}$$

(17)

A modificação de DRC pode agora ser computada da sonoridade após a aplicação da escalagem de AGC. Ao invés de suavizar uma medida da sonoridade antes da aplicação da curva de DRC, pode-se alternativamente aplicar a curva de DRC na sonoridade instantânea e então subsequentemente suavizar a modificação resultante. Isto é similar à técnica anteriormente descrita para suavizar o ganho do DRC tradicional. Além disso, o DRC pode ser aplicado em um modo de múltiplas bandas, significando que a modificação de DRC é uma função da sonoridade específica $N[b,t]$ em cada banda d , ao invés da sonoridade total $L[t]$. No entanto, de modo a manter o balanço espectral médio do áudio original, pode-se aplicar o DRC a cada banda de modo que as modificações resultantes tenham o mesmo efeito médio que resultaria da aplicação de DRC na sonoridade total. Isto pode ser conseguido escalando cada banda pela razão da sonoridade total de longo prazo (após a aplicação da escalagem de AGC) para a sonoridade específica a longo prazo, e utilizando este valor como o argumento para a função de DRC. O resultado é então reescalado pelo inverso da dita razão para produzir a sonoridade específica de saída. Assim a escalagem de DRC em cada banda pode ser computada de acordo com:

$$S_{DRC}[b,t] = \frac{N_{AGC}[b,t]}{S_{AGC}[t]L_{AGC}[t]} F_{DRC} \left\{ \frac{S_{AGC}[t]L_{AGC}[t]}{N_{AGC}[t]} N[b,t] \right\} \quad (18)$$

As modificações de AGC e de DRC podem então ser combinadas para formar a escalagem de sonoridade total por banda:

$$S_{TOT}[b,t] = S_{AGC}[t]S_{DRC}[b,t] \quad (19)$$

Esta escalagem total pode então ser suavizada através do tem-

po independentemente para cada banda com um rápido ataque e uma lenta liberação e o controle de evento aplicado somente à liberação. Idealmente, a suavização é executada no logaritmo da escalagem análogo aos ganhos do DRC tradicional sendo suavizados em sua representação de decibel, apesar disto não ser essencial. Para assegurar que a escalagem total suavizada move-se em sincronismo com a sonoridade específica em cada banda, os modos de ataque e de liberação podem ser determinados através da suavização simultânea da própria sonoridade específica:

$$\bar{S}_{TOT}[b, t] = \exp(\alpha_{TOT}[b, t] \log(\bar{S}_{TOT}[b, t-1]) + (1 - \alpha_{TOT}[b, t]) \log(S_{TOT}[b, t])) \quad (20a)$$

$$\bar{N}[b, t] = \alpha_{TOT}[b, t] \bar{N}[b, t-1] + (1 - \alpha_{TOT}[b, t]) N[b, t]$$

(20b)

onde

$$\alpha_{TOT}[b, t] = \begin{cases} \alpha_{TOTattack} & N[b, t] > \bar{N}[b, t-1] \\ \bar{A}[t] \alpha_{TOTrelease} + (1 - \bar{A}[t]) & N[b, t] \leq \bar{N}[b, t-1] \end{cases}$$

(20c)

10 Finalmente pode-se computar uma sonoridade específica alvo com base na escalagem suavizada aplicada na sonoridade específica original

$$\hat{N}[b, t] = \bar{S}_{TOT}[b, t] N[b, t]$$

(21)

e então resolver para os ganhos $G[b, t]$ que quando aplicados à excitação original resultam em uma sonoridade específica igual ao alvo:

$$\hat{N}[b, t] = \Psi\{G^2[b, t] E[b, t]\}$$

(22)

15 Os ganhos podem ser aplicados a cada banda do banco de filtros utilizado para computar a excitação, e o áudio modificado pode então

ser gerado pela inversão do banco de filtros para produzir um sinal de áudio de domínio de tempo modificado.

Controle de Parâmetro Adicional

Apesar da discussão acima ter focalizado no controle dos parâmetros de ataque e de liberação de AGC e de DRC através de análise de cena auditiva do áudio que está sendo processado, outros parâmetros importantes podem também beneficiar-se de serem controlados através dos resultados de ASA. Por exemplo, o sinal de controle de evento $\bar{A}[t]$ da Equação 12 pode ser utilizado para variar o valor do parâmetro de razão de DRC que é utilizado para ajustar dinamicamente o ganho do áudio. O parâmetro de Razão, similarmente aos parâmetros de tempo de ataque e de liberação, pode contribuir significativamente para os artefatos perceptuais introduzidos por ajustes de ganho dinâmico.

Implementação

A invenção pode ser implementada em hardware ou em software, ou uma combinação de ambos (por exemplo, matrizes de lógica programável). A menos que de outro modo especificado, os algoritmos incluídos como parte da invenção não estão inerentemente relacionados a nenhum computador ou outro aparelho específico. Especificamente, várias máquinas de uso geral podem ser utilizadas com os programas escritos de acordo com os ensinamentos aqui, ou pode ser mais conveniente construir um aparelho mais especializado (por exemplo, circuitos integrados) para executar as etapas de método requeridas. Assim, a invenção pode ser implementada em um ou mais programas de computador que executam em um ou mais sistemas de computador programáveis cada um compreendendo pelo menos um processador, pelo menos um sistema de armazenamento de dados (incluindo uma memória volátil e não-volátil e/ou elementos de armazenamento), pelo menos um dispositivo ou porta de entrada, e pelo menos um dispositivo ou porta de saída. Um código de programa é aplicado aos dados de entrada para executar as funções aqui descritas e gerar as informações de saída. As informações de saídas são aplicadas a um ou mais dispositivo de saída, em um modo conhecido.

Cada tal programa pode ser implementado em qualquer linguagem de computador desejada (incluindo as linguagens de programação de máquina, assembly, ou de procedimentais de alto nível, lógicas, ou orientadas em objeto) para comunicar com um sistema de computador. Em qualquer caso, a linguagem pode ser uma linguagem compilada ou interpretada.

Cada tal programa de computador está de preferência armazenado em ou carregado para um meio ou dispositivo de armazenamento (por exemplo, uma memória ou mídia de estado sólido, ou mídia magnética ou ótica) legível por um computador programado de uso geral ou especial, para configurar e operar o computador quando o meio ou dispositivo de armazenamento é lido pelo sistema de computador para executar os procedimentos aqui descritos. O sistema inventivo pode também ser considerado ser implementado como um meio de armazenamento legível por computador, configurado com um programa de computador, onde o meio de armazenamento assim configurado faz com que um sistema de computador opere em um modo específico e predefinido para executar as funções aqui descritas.

Um número de modalidades da invenção foi descrito. Apesar de tudo, será compreendido que várias modificações podem ser feitas sem afastar-se do espírito e do escopo da invenção. Por exemplo, algumas das etapas aqui descritas podem ser independentes de ordem, e assim podem ser executadas em uma ordem diferente daquela descrita.

Deve ser compreendido que a implementação de outras variações e modificações da invenção e seus vários aspectos serão aparentes para aqueles versados na técnica, e que a invenção não está limitada por estas modalidades específicas descritas. É portanto contemplado cobrir pela presente invenção quaisquer e todas modificações, variações, ou equivalentes que caiam dentro do verdadeiro espírito e escopo dos princípios subjacentes básicos aqui descritos e reivindicados.

Incorporação por Referência

As seguintes patentes, pedidos de patente e publicações estão por meio disto incorporadas por referência, cada uma em sua totalidade.

Processamento de Dinâmica de Áudio

Audio Engineer's Reference Book, editado por Michael Talbot-Smith, 2ª edição. Limiters and Compressors, Alan Tutton, 2-1492-165. Focal Press, Reed Educational and Professional Publishing, Ltd., 1999.

Detectando e Utilizando os Eventos Auditivos

5 Pedido de Patente U.S. S.N. 10/474.387, "High Quality Time-Scaling and Pitch-Scaling of Audio Signals" de Brett Graham Crockett, publicado em 24 de Junho de 2004 como US 2004/0122662 A1.

 Pedido de Patente U.S. S.N. 10/478.398, "Method for Time Aligning Audio Signals Using Characterizations Based on Auditory Events" de
10 Brett G. Crockett et al, publicado em 29 de Julho de 2004 como US 2004/0148159 A1.

 Pedido de Patente U.S. S.N. 10/478.538, "Segmenting Audio Signals Into Auditory Events" de Brett G. Crockett, publicado em 26 de Agosto de 2004 como US 2004/0165730 A1. Aspectos da presente invenção pro-
15 vêm um meio para detectar os eventos auditivos além daqueles descritos no dito pedido de Crockett.

 Pedido de Patente U.S. S.N. 10/478.397, "Comparing Audio Using Characterizations Based on Auditory Events" de Brett G. Crockett et al, publicado em 02 de Setembro de 2004 como US 2004/0172240 A1.

20 Pedido Internacional sob o Tratado de Cooperação de Patentes S.N. PCT/US 05/24630 depositado em 13 de Julho de 2005, intitulado "Method for Combining Audio Signals Using Auditory Scene Analysis", de Michael John Smithers, publicado em 09 de Março de 2006 como WO 2006/026161.

25 Pedido Internacional sob o Tratado de Cooperação de Patentes S.N. PCT/US 2004/016964, depositado em 27 de maio de 2004, intitulado "Method, Apparatus and Computer Program for Calculating and Adjusting the Perceived Loudness of an Audio Signal" de Alan Jeffrey Seefeldt et al, publicado em 23 de Dezembro de 2004 como WO 2004/111994 A2.

30 Pedido Internacional sob o Tratado de Cooperação de Patentes S.N. PCT/US2005/038579, depositado em 25 de Outubro de 2005, intitulado "Calculating and Adjusting the Perceived Loudness and/or the Perceived

Spectral Balance of an Audio Signal" por Alan Jeffrey Seefeldt e publicado como Publicação Internacional Número WO 2006/047600.

"A Method for Characterizing and Identifying Audio Based on Auditory Scene Analysis", por Brett Crockett e Michael Smithers, Audio Engineering Society Convention Paper 6416, 118th Convention, Barcelona, 28-31 de Maio de 2005.

"High Quality Multichannel Time Scaling and Pitch-Shifting using Auditory Scene Analysis", por Brett Crockett, Audio Engineering Society Convention Paper 5948, New York, Outubro de 2003.

"A New Objective Measure of Perceived Loudness" por Alan Seefeldt et al, Audio Engineering Society Convention Paper 6236, San Francisco, 28 de Outubro de 2004.

Handbook for Sound Engineers, The New Audio Cyclopedia, editado por Glen M. Ballou, 2^a Edição. Dynamics, 850-851. Focal Press an imprint of Butterworth-Heinemann, 1998.

Audio Engineer's Reference Book, editado por Michael Talbot-Smith, 2^a Edição, Seção 2.9 ("Limiters and Compressors" por Alan Tutton), pp. 2.149-2.165, Focal Press, Reed Educational and Professional Publishing, Ltd., 1999.

REIVINDICAÇÕES

1. Método de processamento de áudio no qual um processador recebe um canal de entrada e gera um canal de saída que é gerado pela aplicação de modificações de ganho dinâmico no canal de entrada, que
5 compreende:

detectar as mudanças em características de sinal em relação ao tempo no canal de entrada de áudio,

identificar como limites de evento auditivo as mudanças em características de sinal em relação ao tempo no dito canal de entrada, em que
10 um segmento de áudio entre os limites consecutivos constitui um evento auditivo no canal, e

gerar todos ou alguns de um ou mais parâmetros do método de modificação de ganho dinâmico pelo menos parcialmente em resposta a eventos auditivos e/ou o grau de mudança em características de sinal associadas aos ditos limites de evento auditivo.
15

2. Método de acordo com a reivindicação 1, em que um evento auditivo é um segmento de áudio que tende a ser percebido como separado e distinto.

3. Método de acordo com a reivindicação 1 ou reivindicação 2, em que as ditas características de sinal incluem o conteúdo espectral do áudio.
20

4. Método de acordo com a reivindicação 1 ou reivindicação 2, em que as ditas características de sinal incluem a sonoridade perceptual do áudio.

5. Método de acordo com qualquer uma das reivindicações 1 a 4, em que todos ou alguns dos ditos um ou mais parâmetros são gerados pelo menos parcialmente em resposta à presença ou à ausência de um ou mais eventos auditivos.
25

6. Método de acordo com qualquer uma das reivindicações 1 a 4, em que a dita identificação identifica como um limite de evento auditivo uma mudança em características de sinal em relação ao tempo que excede um limite.
30

7. Método de acordo com qualquer uma das reivindicações 1 a 4, em que o dito limite de evento auditivo pode ser modificado por uma função para criar um sinal de controle que é utilizado para modificar os parâmetros de modificação de ganho dinâmico de áudio.

5 8. Método de acordo com qualquer uma das reivindicações 1 a 4, em que todos ou alguns dos ditos um ou mais parâmetros são gerados pelo menos parcialmente em resposta a uma medida continuada do grau de mudança em características de sinal associadas aos ditos limites de eventos auditivos.

10 9. Aparelho adaptado para executar os métodos como definidos em qualquer uma das reivindicações 1 a 8.

10. Programa de computador armazenado em um meio legível por computador, para fazer com que um computador controle o aparelho como definido na reivindicação 9.

15 11. Programa de computador armazenado em um meio legível por computador, para fazer com que um computador execute os métodos de acordo com qualquer uma das reivindicações 1 a 8.

20 12. Método para dividir um sinal de áudio em eventos auditivos, cada um dos quais tende a ser percebido como separado e distinto, que compreende:

calcular a diferença em conteúdo espectral entre os sucessivos blocos de tempo do dito sinal de áudio, em que a diferença é calculada pela comparação da diferença em sonoridade específica entre os sucessivos blocos de tempo, em que a sonoridade específica é uma medida de sonoridade perceptual como uma função de frequência e tempo, e

25 identificar um limite de evento auditivo como o limite entre os sucessivos blocos de tempo quando a diferença em conteúdo espectral entre tais sucessivos blocos de tempo excede um limite.

30 13. Método de acordo com a reivindicação 12, em que o dito sinal de áudio é representado por uma sequência de tempo discreta $x[n]$ que foi mostrada de uma fonte de áudio a uma frequência de amostragem f_s e a diferença é calculada pela comparação da diferença em sonoridade especí-

fica $N[b,t]$ através de bandas de frequência b entre os sucessivos blocos de tempo t .

14. Método de acordo com a reivindicação 13, em que a diferença em conteúdo espectral entre os sucessivos blocos de tempo do sinal de áudio é calculada de acordo com

$$D[t] = \sum_b |N_{NORM}[b,t] - N_{NORM}[b,t-1]|$$

onde

$$N_{NORM}[b,t] = \frac{N[b,t]}{\max_b \{N[b,t]\}}$$

15. Método de acordo com a reivindicação 13, em que a diferença em conteúdo espectral entre os sucessivos blocos de tempo do sinal de áudio é calculada de acordo com

$$D[t] = \sum_b |N_{NORM}[b,t] - N_{NORM}[b,t-1]|$$

10 onde

$$N_{NORM}[b,t] = \frac{N[b,t]}{\text{avg}_b \{N[b,t]\}}$$

16. Aparelho adaptado para executar os métodos como definidos em qualquer uma das reivindicações 12 a 15.

17. Programa de computador, armazenado em um meio legível por computador, para fazer com que um computador controle o aparelho como definido na reivindicação 16.

18. Programa de computador armazenado em um meio legível por computador, para fazer com que um computador execute os métodos como definidos em qualquer uma das reivindicações 12 a 15.

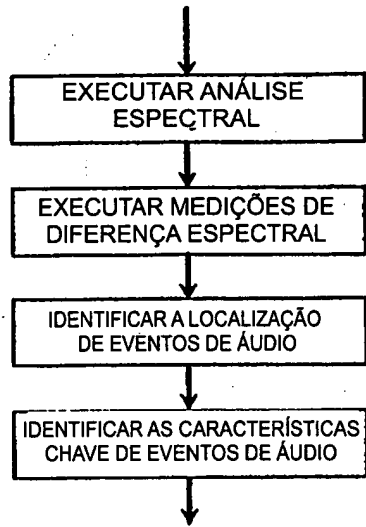


FIG. 1

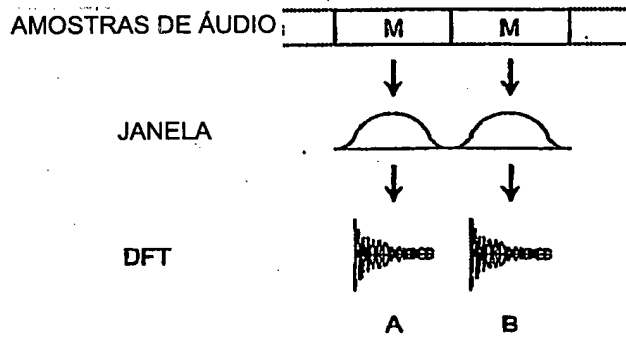


FIG. 2

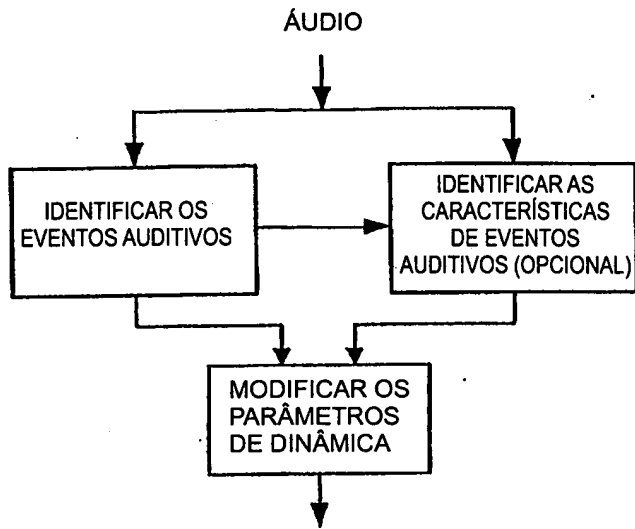


FIG. 3

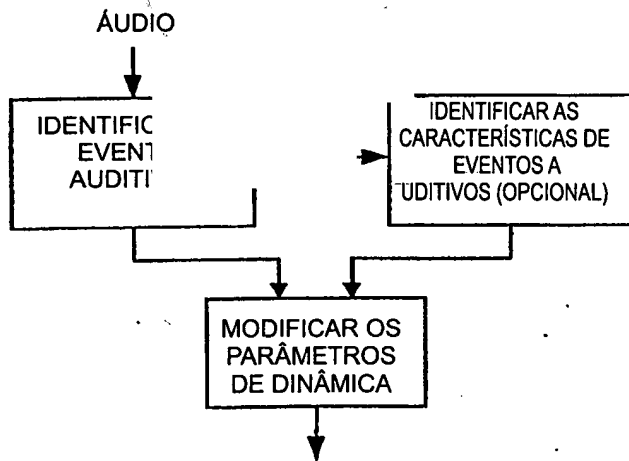


FIG. 4

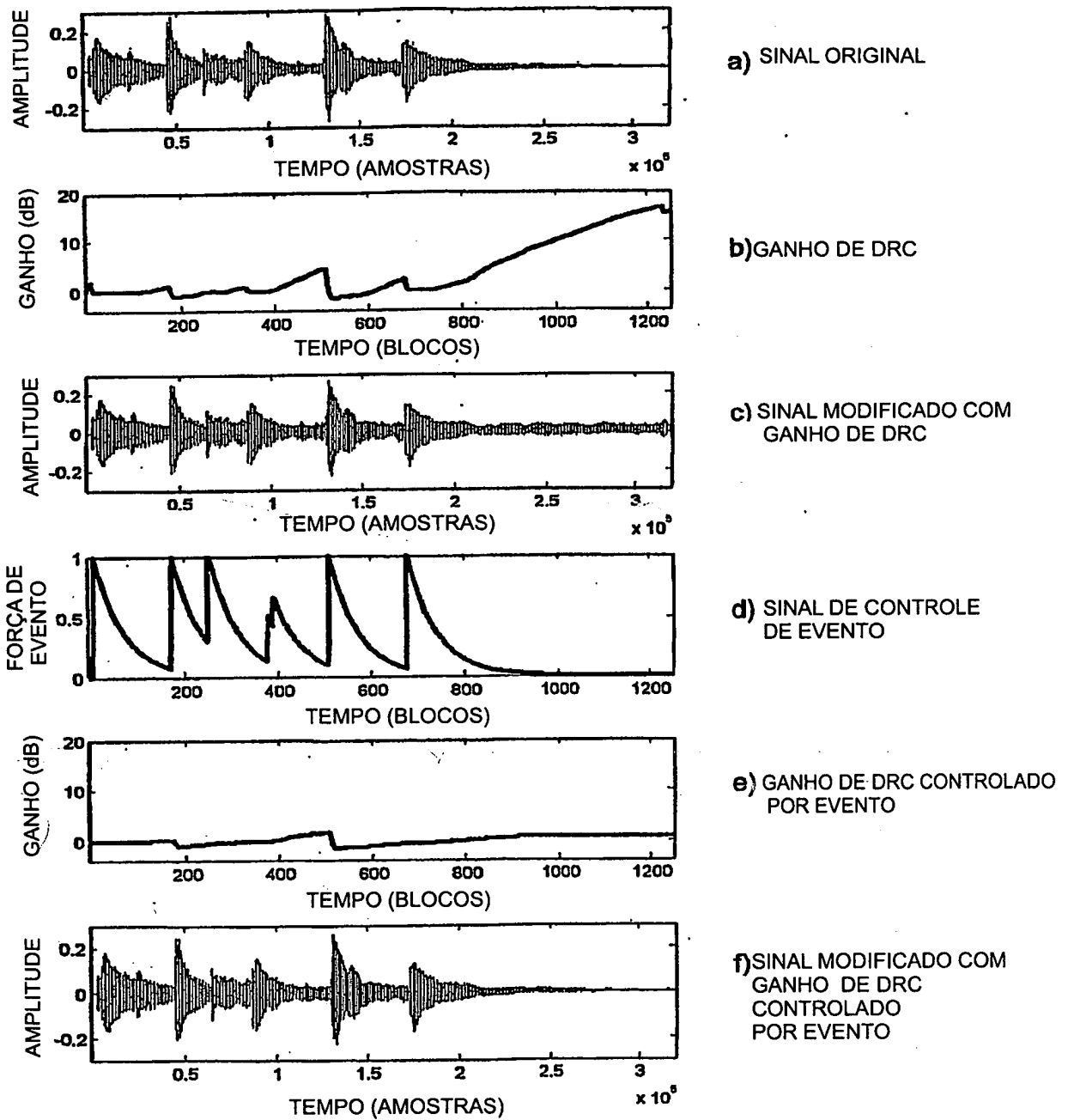


FIG. 9

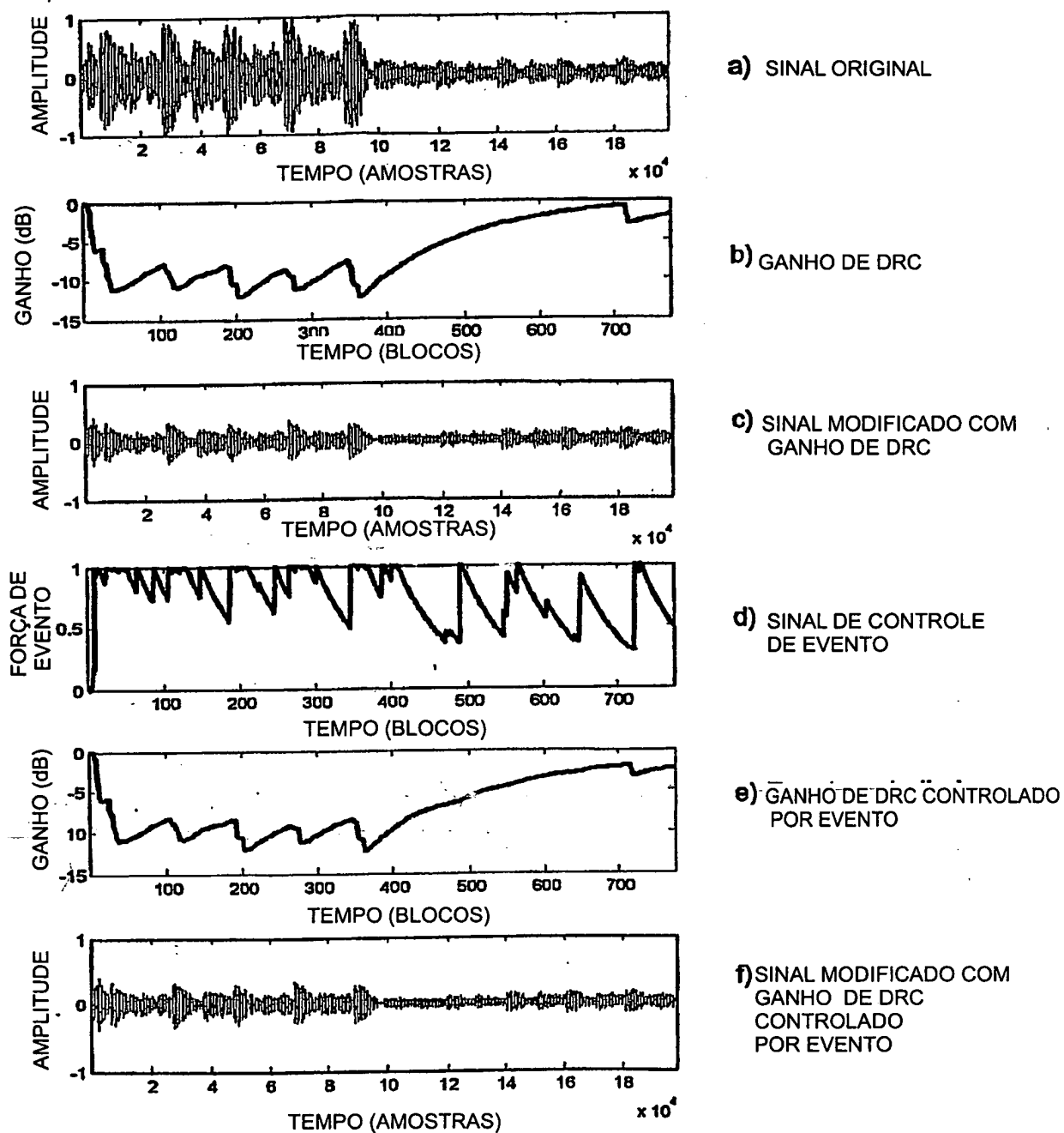
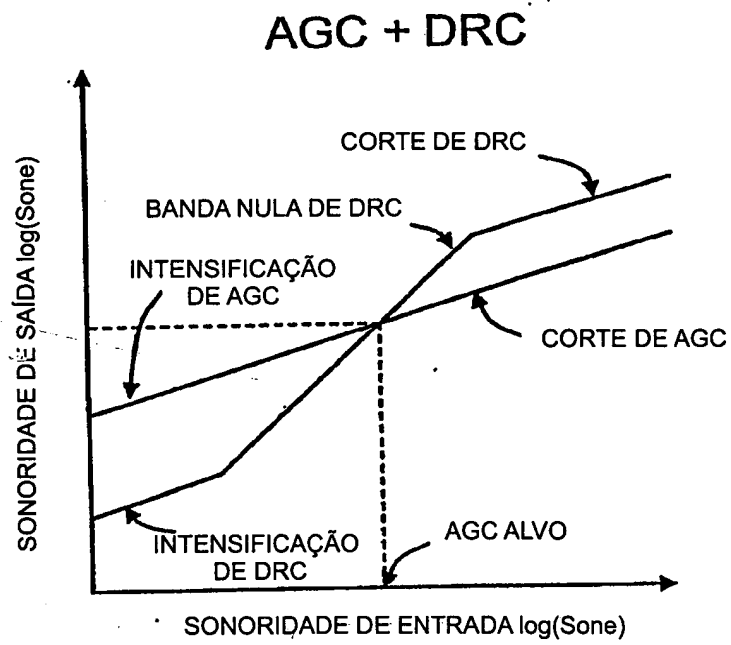


FIG. 10

**FIG. 11**

PI0711063-4

RESUMO

Patente de Invenção: **"CONTROLE DE GANHO DE ÁUDIO UTILIZANDO DETECÇÃO DE EVENTO AUDITIVO BASEADÔ EM SONORIDADE ESPECÍFICA"**.

5 A presente invenção refere-se a modificações de ganho dinâmico aplicadas a um sinal de áudio pelo menos parcialmente em resposta a eventos auditivos e/ou o grau de mudança em características de sinal associadas aos ditos limites de evento auditivo. Em outro aspecto, um sinal de áudio é dividido em eventos auditivos pela comparação da diferença em sonoridade específica entre os sucessivos blocos de tempo do sinal de áudio.

10