



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

(21) **PI 0911456-4 A2**

(22) Data de Depósito: 17/04/2009  
(43) Data da Publicação: 07/05/2013  
(RPI 2209)



(51) *Int.Cl.:*  
G10L 21/02  
H04S 3/00  
G10L 11/02

(54) **Título:** MÉTODO E APARELHO PARA MANTER AUDIBILIDADE DE FALA EM ÁUDIO DE MÚLTIPLOS CANAIS COM IMPACTOS MÍNIMO EM EXPERIÊNCIA ENVOLVENTE

(30) **Prioridade Unionista:** 18/04/2008 US 61/046,271

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

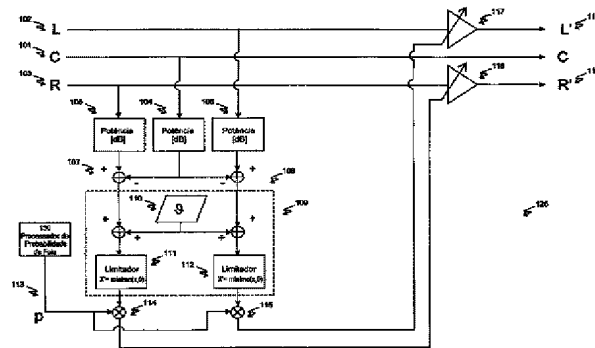
(72) **Inventor(es):** Hannes Muesch

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

(86) **Pedido Internacional:** PCT US2009040900 de 17/04/2009

(87) **Publicação Internacional:** WO 2010/011377de 28/01/2010

(57) **Resumo:** MÉTODO E APARELHO PARA MANTER AUDIBILIDADE DE FALA EM ÁUDIO DE MÚLTIPLOS CANAIS COM IMPACTO MÍNIMO EM EXPERIÊNCIA ENVOLVENTE. A presente invenção refere-se a um método de aperfeiçoar audibilidade de fala em um sinal de áudio de múltiplos canais. O método inclui comparar uma primeira característica e uma segunda característica do sinal de áudio de múltiplos canais gerar um fator de atenuação. A primeira característica corresponde a um primeiro canal do sinal de áudio de múltiplos canais que contém áudio de fala e não fala, e a segunda característica corresponde a um segundo canal do sinal de áudio de múltiplos canais que contém predominantemente áudio de não fala. O método inclui adicionalmente ajustar o fator de atenuação de acordo com um valor de probabilidade de fala para gerar um fator de atenuação ajustado. O método inclui adicionalmente atenuar o segundo canal usando o fator de atenuação ajustado.



Relatório Descritivo da Patente de Invenção para **"MÉTODO E APARELHO PARA MANTER AUDIBILIDADE DE FALA EM ÁUDIO DE MÚLTIPLOS CANAIS COM IMPACTO MÍNIMO EM EXPERIÊNCIA ENVOLVENTE"**.

5 Referência Cruzada a Pedidos Relacionados

Este pedido reivindica o benefício de prioridade do Pedido de Patente Provisório Estados Unidos Nº 61/046.271, depositado em 18 de abril de 2008, incorporado na sua totalidade neste pedido pela referência.

Antecedentes

10 A presente invenção refere-se a processamento de sinal de áudio em geral e a aperfeiçoar clareza de diálogo e narrativa em áudio de entretenimento envolvente em particular.

A não ser que indicado de outro modo neste documento, as abordagens descritas nesta seção não são de técnica anterior para as reivindicações neste pedido e não são admitidas para serem de técnica anterior por inclusão nesta seção.

20 Áudio de entretenimento moderno com múltiplos canais simultâneos de áudio (som envolvente) fornece audiências com ambientes de som realísticos e imersivos de imenso valor de entretenimento. Em tais ambientes muitos elementos de som tais como diálogo, música e efeitos são apresentados simultaneamente e competem pela atenção do ouvinte. Para algumas pessoas da audiência - especialmente aquelas com capacidades sensoriais auditivas reduzidas ou processamento cognitivo reduzido - diálogo e narrativa podem ser difíceis de entender durante partes do programa onde elementos de som concorrendo estrondosamente estão presentes. Durante essas passagens estes ouvintes seriam beneficiados se o nível dos sons concorrentes fosse abaixado.

30 O reconhecimento de que música e efeitos podem sobrepujar diálogo não é novo e diversos métodos para solucionar a situação têm sido sugeridos. Entretanto, tal como será delineado a seguir, os métodos sugeridos são incompatíveis com a prática de difusão corrente, ou aplicam um preço desnecessariamente alto no experimento de entretenimento total, ou

ambos.

É uma convenção comumente adotada na produção de áudio envolvente para filme e televisão colocar a maior parte de diálogo e narrativa somente em um canal (o canal central, também referido como o canal de fala). Música, sons ambientes e efeitos sonoros são tipicamente mixados tanto no canal de fala quanto em todos os canais restantes (por exemplo, Left [L], Right [R], Envolvente Esquerdo [ls] e Envolvente Direito [rs], também referidos como os canais de não-fala). Como resultado, o canal de fala carrega a maioria da fala e uma quantidade significativa do áudio de não-fala contido no programa de áudio, enquanto que os canais de não fala carregam predominantemente áudio de não fala, mas também podem carregar uma pequena quantidade de fala. Uma abordagem simples para ajudar na percepção de diálogo e narrativa nestas mixagens convencionais é reduzir permanentemente o nível de todos os canais de não fala em relação ao nível do canal de fala, por exemplo, por 6 dB. Esta abordagem é simples e efetiva e é praticada atualmente (por exemplo, Clareza de Diálogo SRS [Sistema de Recuperação de Som] ou equações de submixagens modificadas em decodificadores envolventes). Entretanto, ela sofre de pelo menos uma desvantagem: a atenuação constante dos canais de não fala pode diminuir o nível de sons ambientes tranquilos que não interferem com recepção de fala para o ponto onde eles não podem mais ser ouvidos. Ao atenuar sons ambientes não interferindo o balanço estético do programa é alterado sem qualquer benefício concomitante para entendimento de fala.

Uma solução alternativa está descrita em uma série de patentes (Patente U.S. Nº 7.266.501, Patente U.S. Nº 6.772.127, Patente U.S. No 6.912.501 e Patente U.S. Nº 6.650.755) para Vaudrey e Saunders. Tal como entendido, sua abordagem envolve modificar a produção e distribuição de conteúdo. De acordo com esse arranjo, o consumidor recebe dois sinais de áudio separados. O primeiro destes sinais compreende o áudio de "Conteúdo Primário". Em muitos casos este sinal será dominado por fala, mas, se o produtor de conteúdo desejar, poderá conter outros tipos de sinal igualmente. O segundo sinal compreende o áudio de "Conteúdo Secundário", o qual é

composto de todos os elementos de sons restantes. Ao usuário é dado controle sobre os níveis relativos destes dois sinais, ao ajustar manualmente o nível de cada sinal ou ao manter automaticamente uma relação de potência selecionada pelo usuário. Embora este arranjo possa limitar a atenuação desnecessária de sons ambientes não interferindo, sua implementação mais difundida é obstruída por sua incompatibilidade com métodos de produção e distribuição estabelecidos.

Um outro exemplo de um método para gerenciar os níveis relativos de áudio de fala e não fala foi proposto por Bennett na Publicação de Pedido U.S. Nº 20070027682.

Todos os exemplos da técnica anterior compartilham a limitação de não fornecer quaisquer meios para minimizar o efeito que o aperfeiçoamento de diálogo tem sobre a experiência de ouvir pretendida pelo criador de conteúdo, entre outras deficiências. Portanto, é o objetivo da presente invenção fornecer um meio de limitar o nível de canais de áudio de não fala em um programa de entretenimento de múltiplos canais mixados convencionalmente de maneira que fala permanece compreensível enquanto mantendo também a audibilidade dos componentes de áudio de não fala.

Assim, existe uma necessidade de modos aperfeiçoados de manter audibilidade de fala. A presente invenção resolve estes e outros problemas ao fornecer um aparelho e método de aperfeiçoar audibilidade de fala em um sinal de áudio de múltiplos canais.

### Sumário

Modalidades da presente invenção aperfeiçoam audibilidade de fala. Em uma modalidade a presente invenção inclui um método de aperfeiçoar audibilidade de fala em um sinal de áudio de múltiplos canais. O método inclui comparar uma primeira característica e uma segunda característica do sinal de áudio de múltiplos canais para gerar um fator de atenuação. A primeira característica corresponde a um primeiro canal do sinal de áudio de múltiplos canais que contém áudio de fala e não fala, e a segunda característica corresponde a um segundo canal do sinal de áudio de múltiplos canais que contém predominantemente áudio de não fala. O método inclui adi-

cionalmente ajustar o fator de atenuação de acordo com um valor de probabilidade de fala para gerar um fator de atenuação ajustado. O método inclui adicionalmente atenuar o segundo canal usando o fator de atenuação ajustado.

5 Um primeiro aspecto da invenção é baseado na observação de que o canal de fala de um programa de entretenimento típico carrega um sinal de não fala por uma parte substancial da duração de programa. Consequentemente de acordo com este primeiro aspecto da invenção, mascaramento de áudio de fala por áudio de não fala pode ser controlado por meio  
10 de (a) determinar a atenuação de um sinal em um canal de não fala necessária para limitar a razão da potência de sinal no canal de não fala para a potência de sinal no canal de fala para não exceder um limiar predeterminado e (b) escalar a atenuação por um fator que seja relacionado de forma uniforme com a probabilidade do sinal no canal de fala ser fala, e (c) aplicar a  
15 atenuação escalada.

Um segundo aspecto da invenção é baseado na observação de que a razão entre a potência do sinal de fala e a potência do sinal de mascaramento é um preditor inferior de inteligibilidade de fala. Consequentemente, de acordo com este segundo aspecto da invenção, a atenuação do sinal no  
20 canal de não fala que é necessária para manter um nível predeterminado de inteligibilidade é calculada ao predizer a inteligibilidade do sinal de fala na presença dos sinais de não fala com um modelo de predição de inteligibilidade baseado em psicoacústica.

Um terceiro aspecto da invenção é baseado nas observações de  
25 que, se atenuação for permitida para variar através da frequência, (a) um dado nível de inteligibilidade pode ser alcançado com uma variedade de padrões de atenuação, e (b) diferentes padrões de atenuação podem produzir diferentes níveis de sonoridade ou ênfase do áudio de não fala. Consequentemente de acordo com este terceiro aspecto da invenção, mascaramento  
30 de áudio de fala por áudio de não fala é controlado ao descobrir o padrão de atenuação que maximiza sonoridade ou alguma outra medida de ênfase do áudio de não fala sob a restrição de que um nível predeterminado de inteli-

bilidade de fala predita é alcançado.

As modalidades da presente invenção podem ser executadas como um método ou processo. Os métodos podem ser implementados por meio de conjunto de circuitos eletrônicos, tais como hardware ou software ou  
5 uma combinação dos mesmos. O conjunto de circuitos usado para implementar o processo pode ser um conjunto de circuitos dedicado (que executa somente uma tarefa específica) ou conjunto de circuitos geral (que é programado para executar uma ou mais tarefas específicas).

A descrição detalhada a seguir e os desenhos anexos fornecem  
10 um melhor entendimento da natureza e vantagens da presente invenção.

#### Breve descrição dos Desenhos

A figura 1 ilustra um processador de sinal de acordo com uma modalidade da presente invenção.

A figura 2 ilustra um processador de sinal de acordo com uma  
15 outra modalidade da presente invenção.

A figura 3 ilustra um processador de sinal de acordo com uma outra modalidade da presente invenção.

As figuras 4A-4B são diagramas de blocos ilustrando variações adicionais das modalidades das figuras 1-3.

#### 20 Descrição Detalhada

São descritas neste documento técnicas para manter audibilidade de fala. Na descrição a seguir, para propósitos de explicação, inúmeros exemplos e detalhes específicos estão expostos a fim de fornecer um completo entendimento da presente invenção. Estará evidente, entretanto, para  
25 os versados na técnica que a presente invenção tal como definida pelas reivindicações pode incluir alguns ou todos os recursos nestes exemplos sozinhos ou em combinação com outros recursos descritos a seguir, e pode incluir adicionalmente modificações e equivalências dos recursos e conceitos descritos neste documento.

30 Vários métodos e processos são descritos a seguir. Que eles são descritos em uma certa ordem é principalmente para facilidade de apresentação. É para ser entendido que etapas particulares podem ser executa-

das em outras ordens ou em paralelo tal como desejado de acordo com várias implementações. Quando uma etapa particular deve preceder ou seguir uma outra, isto será salientado especificamente quando não evidente a partir do contexto.

5                   O princípio do primeiro aspecto da invenção está ilustrado na figura 1. Referindo-se agora à figura 1, um sinal de múltiplos canais consistindo em um canal de fala (101) e dois canais de não fala (102 e 103) é recebido. A potência dos sinais em cada um destes canais é medida com um banco de estimadores de potência (104, 105 e 106) e expressada em uma escala  
10                   logarítmica [dB]. Estes estimadores de potência podem conter um mecanismo de uniformização, tal como um integrador com fuga, de maneira que o nível de potência medido reflete o nível de potência determinado pela média sobre a duração de uma sentença ou uma passagem total. O nível de potência do sinal no canal de fala é subtraído do nível de potência em cada um  
15                   dos canais de não fala (pelos somadores 107 e 108) para dar uma medida da diferença de nível de potência entre os dois tipos de sinais. O circuito de comparação 109 determina para cada canal de não fala o número de dB pelo qual o canal de não fala deve ser atenuado a fim de seu nível de potência permanecer pelo menos  $v$  dB abaixo do nível de potência do sinal no canal  
20                   de fala (o símbolo " $v$ " indica uma variável e também pode ser referido como a letra grega teta ( $\theta$ ) manuscrita). De acordo com uma modalidade, uma implementação disto é adicionar o valor limiar  $v$  (armazenado pelo circuito 110) à diferença de nível de potência (este resultado intermediário é referido como a margem) e limitar o resultado para ser igual ou menor que zero (pelos  
25                   limitadores 111 e 112). O resultado é o ganho (ou atenuação negada) em dB que deve ser aplicado aos canais de não fala para manter seu nível de potência  $v$  dB abaixo do nível de potência do canal de fala. Um valor adequado para  $v$  é 15 dB. O valor de  $v$  pode ser ajustado tal como desejado em outras modalidades.

30                   Por causa de existir uma relação exclusiva entre uma medida expressada em uma escala logarítmica (dB) e essa mesma medida expressada em uma escala linear, um circuito que é equivalente à figura 1 pode ser

construído onde potência, ganho e limiar são todos expressados em uma escala linear. Nessa implementação todas as diferenças de nível são substituídas por razões das medidas lineares. Implementações alternativas podem substituir a medida de potência por medidas que estejam relacionadas com intensidade de sinal, tal como o valor absoluto do sinal.

Um recurso digno de nota do primeiro aspecto da invenção é escalar o ganho assim derivado por um valor de forma uniforme relacionado com a probabilidade de o sinal no canal de fala ser fala de fato. Referindo-se ainda à figura 1, um sinal de controle (113) é recebido e multiplicado pelos ganhos (pelos multiplicadores 114 e 115). Os ganhos escalados são então aplicados aos canais de não fala correspondentes (pelos amplificadores 116 e 117) para produzir os sinais modificados L' e R' (118 e 119). O sinal de controle (113) tipicamente será uma medida derivada automaticamente da probabilidade de o sinal no canal de fala ser fala. Vários métodos de determinar automaticamente a probabilidade de um sinal ser um sinal de fala podem ser usados. De acordo com uma modalidade, um processador de probabilidade de fala 130 gera o valor de probabilidade de fala p (113) a partir da informação no canal C 101. Um exemplo de um mecanismo como este está descrito por Robinson e Vinton em "Automated Speech/Other Discrimination for Loudness Monitoring" (Audio Engineering Society, número pré-impresso 6437 da Convenção 118 de maio de 2005). Alternativamente, o sinal de controle (113) pode ser criado manualmente, por exemplo, pelo criador de conteúdo e transmitido ao lado do sinal de áudio para o usuário final.

Os versados na técnica reconhecerão facilmente como o arranjo pode ser estendido para qualquer número de canais de entrada.

O princípio do segundo aspecto da invenção está ilustrado na figura 2. Referindo-se agora à figura 2, um sinal de múltiplos canais consistindo em um canal de fala (101) e dois canais de não fala (102 e 103) é recebido. A potência dos sinais em cada um destes canais é medida com um banco de estimadores de potência (201, 202 e 203). Ao contrário de suas contrapartes na figura 1, estes estimadores de potência mediram a distribuição da potência de sinal através da frequência, resultando em um espectro de

potência em vez de um único número. A resolução espectral do espectro de potência idealmente casa com a resolução espectral do modelo de predição de inteligibilidade (205 e 206, ainda não discutido).

Os espectros de potência são alimentados para o circuito de  
5 comparação 204. O propósito deste bloco é determinar a atenuação a ser aplicada a cada canal de não fala para assegurar que o sinal no canal de não fala não reduz a inteligibilidade do sinal no canal de fala para ser menos que um critério predeterminado. Esta funcionalidade é alcançada ao empregar um circuito de predição de inteligibilidade (205 e 206) que prediz  
10 inteligibilidade de fala a partir dos espectros de potência do sinal de fala (201) e dos sinais de não fala (202 e 203). Os circuitos de predição de inteligibilidade de 205 e 206 podem implementar um modelo de predição de inteligibilidade adequado de acordo com opções de projeto e trocas compensatórias. Exemplos são o Índice de Inteligibilidade de Fala tal como especificado na  
15 ANSI S3.5-1997 ("Methods for Calculation of the Speech Index") e o modelo de Sensibilidade de Reconhecimento de Fala de Muesch e Buus ("Using statistical decision theory to predict speech intelligibility. I. Model structure" *Jornal da Acoustical Society of America*, 2001, Vol 109, p 2896-2909). É claro que a saída do modelo de predição de inteligibilidade não tem significado  
20 quando o sinal no canal de fala é alguma coisa a não ser fala. Apesar disto, no que se segue a saída do modelo de predição de inteligibilidade será referida como a inteligibilidade de fala predita. O erro percebido será considerado no processamento subsequente ao escalar a saída de valores de ganho do circuito de comparação 204 com um parâmetro que está relacionado com  
25 a probabilidade do sinal ser fala (113, ainda não discutido).

Os modelos de predição de inteligibilidade têm em comum que eles predizem inteligibilidade de fala aumentada ou inalterada como o resultado de abaixamento do nível do sinal de não fala. Continuando no fluxo de processo da figura 2, os circuitos de comparação 207 e 208 comparam a  
30 inteligibilidade predita com um valor de critério. Se o nível do sinal de não fala for baixo de maneira que a inteligibilidade predita excede o critério, o parâmetro de ganho, o qual é inicializado para 0 dB, é recuperado do circuito

209 ou 210 e fornecido para os circuitos 211 e 212 como a saída do circuito de comparação 204. Se o critério não for satisfeito, o parâmetro de ganho é diminuído por uma quantidade fixada e a predição de inteligibilidade é repetida. Um tamanho de incremento adequado para diminuir o ganho é 1 dB. A  
5 iteração tal como descrita exatamente continua até que a inteligibilidade predita satisfaça ou exceda o valor de critério. Certamente é possível que o sinal no canal de fala seja de tal maneira que a inteligibilidade de critério não possa ser alcançada mesmo na falta de um sinal no canal de não fala. Um exemplo de uma situação como esta é um sinal de fala de nível muito inferior  
10 ou com largura de banda seriamente restringida. Se isso acontecer será alcançado um ponto onde qualquer redução adicional do ganho aplicado ao canal de não fala não afetará a inteligibilidade de fala predita e o critério nunca será satisfeito. Em uma condição como esta, o laço formado por (205, 206), (207, 208) e (209, 210) continua indefinidamente, e lógica adicional  
15 (não mostrada) pode ser aplicada para romper o laço. Um exemplo particularmente simples de tal lógica é contar o número de iterações e sair do laço uma vez que um número predeterminado de iterações tenha sido excedido.

Continuando no fluxo de processo da figura 2, um sinal de controle p (113) é recebido e multiplicado pelos ganhos (pelos multiplicadores  
20 114 e 115). O sinal de controle (113) tipicamente será uma medida derivada automaticamente da probabilidade de o sinal no canal de fala ser fala. Métodos de determinar automaticamente a probabilidade de um sinal ser um sinal de fala são conhecidos por si e foram discutidos no contexto da figura 1 (vide o processador de probabilidade de fala 130). Os ganhos escalados são en-  
25 tão aplicados aos seus canais de não fala correspondentes (pelos amplificadores 116 e 117) para produzir os sinais modificados R' e L' (118 e 119).

O princípio do terceiro aspecto da invenção está ilustrado na figura 3. Referindo-se agora à figura 3, um sinal de múltiplos canais consistindo em um canal de fala (101) e dois canais de não fala (102 e 103) é recebido.  
30 Cada um dos três sinais é dividido nos seus componentes espectrais (pelos bancos de filtros 301, 302 e 303). A análise espectral pode ser alcançada com um banco de filtros de N canais de domínio de tempo. De acordo

com uma modalidade, o banco de filtros divide a faixa de frequências em bandas de 1/3 de oitava ou se parece com a filtragem suposta para ocorrer no ouvido interno humano. O fato de que o sinal agora consiste em N subsinais está ilustrado pelo uso de linhas em negrito. O processo da figura 3 pode ser reconhecido como um processo de linha secundária lateral. Seguindo o caminho de sinal, cada um dos N subsinais que formam os canais de não fala é escalado por um elemento de um conjunto de N valores de ganho (pelos amplificadores 116 e 117). A derivação destes valores de ganho será descrita mais tarde. A seguir, os subsinais escalados são recombina-  
5 em um único sinal de áudio. Isto pode ser feito por meio de soma simples (pelos circuitos de soma 313 e 314). Alternativamente, um banco de filtros de síntese que é casado com o banco de filtros de análise pode ser usado. Este processo resulta nos sinais de não fala modificados R' e L' (118 e 119).  
10

Descrevendo agora o caminho de linha secundária lateral do processo da figura 3, cada saída de banco de filtros é tornada disponível para um banco correspondente de N estimadores de potência (304, 305 e 306). Os espectros de potência resultantes servem como entradas para um circuito de otimização (307 e 308) que tem como saída um vetor de ganho N dimensional. A otimização emprega tanto um circuito de predição de inteligibilidade (309 e 310) quanto um circuito de cálculo de sonoridade (311 e 312) para descobrir o vetor de ganho que maximiza sonoridade do canal de não fala enquanto mantendo um nível predeterminado de inteligibilidade predita do sinal de fala. Modelos adequados para prever inteligibilidade foram discutidos em conexão com a figura 2. Os circuitos de cálculo de sonoridade  
20 311 e 312 podem implementar um modelo de predição de sonoridade adequado de acordo com opções de projeto e trocas compensatórias. Exemplos de modelos adequados são o Padrão Nacional Americano ANSI S3.4-2007 "Procedure for the Computation of Loudness of Steady Sounds" e o padrão alemão DIN 45631 "Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum".  
25  
30

Dependendo dos recursos computacionais disponíveis e das restrições impostas, a forma e complexidade dos circuitos de otimização

(307, 308) pode variar muito. De acordo com uma modalidade uma otimização restringida multidimensional iterativa de  $N$  parâmetros livres é usada. Cada parâmetro representa o ganho aplicado a uma das bandas de frequência do canal de não fala. Técnicas padrões, tais como seguir o gradiente mais alto no espaço de pesquisa  $N$  dimensional, podem ser aplicadas para descobrir o máximo. Em uma outra modalidade, uma abordagem exigindo menos de forma computacional restringe as funções de ganho versus frequência para serem elementos de um pequeno conjunto de possíveis funções de ganho versus frequência, tal como um conjunto de diferentes gradientes espectrais ou filtros de prateleira. Com esta restrição adicional o problema de otimização pode ser reduzido para um pequeno número de otimizações unidimensionais. Também em uma outra modalidade uma pesquisa exaustiva é feita sobre um conjunto muito pequeno de possíveis funções de ganho. Esta última abordagem pode ser particularmente desejável em aplicações em tempo real onde uma carga computacional constante e velocidade de pesquisa são desejadas.

Os versados na técnica reconhecerão facilmente restrições adicionais que podem ser impostas à otimização de acordo com modalidades adicionais da presente invenção. Um exemplo é restringir a sonoridade do canal de não fala modificado para não ser maior que a sonoridade antes da modificação. Um outro exemplo é impor um limite para as diferenças de ganho entre bandas de frequência adjacentes a fim de limitar o potencial para serrilhado temporal no banco de filtros de reconstrução (313, 314) ou para reduzir a possibilidade para modificações de timbre censuráveis. Restrições desejáveis dependem tanto da implementação técnica do banco de filtros quanto da troca compensatória escolhida entre melhoramento de inteligibilidade e modificação de timbre. Para clareza de ilustração, estas restrições estão omitidas na figura 3.

Continuando no fluxo de processo da figura 3, um sinal de controle  $p$  (113) é recebido e multiplicado com as funções de ganho (pelos multiplicadores 114 e 115). O sinal de controle (113) tipicamente será uma medida derivada automaticamente da probabilidade de o sinal no canal de fala

ser fala. Métodos adequados para calcular automaticamente a probabilidade de um sinal ser fala foram discutidos em conexão com a figura 1 (vide o processador de probabilidade de fala 130). As funções de ganho escalado são então aplicadas aos seus canais de não fala correspondentes (pelos amplificadores 116 e 117), tal como descrito anteriormente.

As figuras 4A e 4B são diagramas de blocos ilustrando variações dos aspectos mostrados nas figuras 1-3. Além do mais, os versados na técnica reconhecerão diversos modos de combinar os elementos da invenção descritos nas figuras 1 a 3.

A figura 4A mostra que o arranjo da figura 1 também pode ser aplicado para uma ou mais sub-bandas de frequência de L, C e R. Especificamente, cada um dos sinais L, C e R pode ser passado através de um banco de filtros (441, 442 e 443), produzindo três conjuntos de n sub-bandas:  $\{L_1, L_2, \dots, L_n\}$ ,  $\{C_1, C_2, \dots, C_n\}$  e  $\{R_1, R_2, \dots, R_n\}$ . Sub-bandas de comparação são passadas para n instâncias do circuito 125 ilustrado na figura 1, e os subsinais processados são recombinados (pelos circuitos de soma 451 e 452). Um valor limiar separado  $v_n$  pode ser selecionado para cada sub-banda. Uma boa escolha é um conjunto onde  $v_n$  é proporcional ao número médio de deixas de fala transportadas na região de frequência correspondente; isto é, bandas nos extremos do espectro de frequência são designadas com limiares mais baixos que bandas correspondendo às frequências de fala dominantes. Esta implementação da invenção oferece uma troca compensatória muito boa entre complexidade computacional e desempenho.

A figura 4B mostra uma outra variação. Por exemplo, para reduzir a carga computacional, um sinal sonoro envolvente típico com cinco canais (C, L, R, ls e rs) pode ser aprimorado ao processar os sinais L e R de acordo com o circuito 325 mostrado na figura 3, e os sinais ls e rs, os quais tipicamente são menos potentes que os sinais L e R, de acordo com o circuito 125 mostrado na figura 1.

Na descrição anterior, os termos e expressões "fala" (ou áudio de fala ou canal de fala ou sinal de fala) e " não fala" (ou áudio de não fala ou canal de não fala ou sinal de não fala) são usados. Os versados na técnica

ca reconhecerão que estes termos e expressões são usados mais para diferenciar uns dos outros e menos para serem descritores absolutos do conteúdo dos canais. Por exemplo, em uma cena de restaurante em um filme, o canal de fala pode conter predominantemente o diálogo em uma mesa e os canais de não fala podem conter o diálogo em outras mesas (consequentemente, ambos contêm "fala" tal como um leigo usa o termo). Também é o diálogo em outras mesas que certas modalidades da presente invenção são direcionadas para atenuar.

### Implementação

10 A invenção pode ser implementada em hardware ou software, ou em uma combinação de ambos (por exemplo, matrizes lógicas programáveis). A não ser que especificado de outro modo, os algoritmos incluídos como parte da invenção não estão relacionados inerentemente a qualquer computador ou outro aparelho particular. Em particular, várias máquinas de uso geral podem ser usadas com programas gravados de acordo com os preceitos neste documento, ou pode ser mais conveniente construir aparelho mais especializado (por exemplo, circuitos integrados) para executar as etapas de método exigidas. Assim, a invenção pode ser implementada em um ou mais programas de computador executando em um ou mais sistemas de computador programável, cada um compreendendo pelo menos um processador, pelo menos um sistema de armazenamento de dados (incluindo 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. Código de programa é aplicado para introduzir dados para executar as funções descritas neste documento e gerar informação de saída. A informação de saída é aplicada a um ou mais dispositivos de saída, em modo conhecido.

Cada tal programa pode ser implementado em qualquer linguagem de computador desejada (incluindo linguagens de máquina, montagem, ou de alto nível processual, lógica, ou de programação orientada a objeto) para se comunicar com um sistema de computador. Em qualquer caso, a linguagem pode ser uma linguagem compilada ou interpretada.

Cada tal programa de computador preferivelmente é armazenado ou transferido para uma mídia ou dispositivo de armazenamento (por exemplo, memória ou mídia de estado sólido, ou mídia magnética ou ótica) legível por um computador programável de uso geral ou especial, para configurar e operar o computador quando a mídia ou dispositivo de armazenamento é lido pelo sistema de computador para executar os procedimentos descritos neste documento. O sistema inventivo também pode ser considerado para ser implementado como uma mídia de armazenamento legível por computador, configurada com um programa de computador, onde a mídia de armazenamento assim configurada induz um sistema de computador para operar em um modo específico e predefinido para executar as funções descritas neste documento.

A descrição anterior ilustra várias modalidades da presente invenção juntamente com exemplos de como aspectos da presente invenção podem ser implementados. Os exemplos e modalidades anteriores não devem ser supostos como sendo as únicas modalidades, e são apresentados para ilustrar a flexibilidade e vantagens da presente invenção tal como definida pelas reivindicações a seguir. Com base na descrição anterior e nas reivindicações a seguir, outros arranjos, modalidades, implementações e equivalências estarão evidentes para os versados na técnica e poderão ser empregados sem divergir do espírito e escopo da invenção tal como definido pelas reivindicações.

## REIVINDICAÇÕES

1. Método de aperfeiçoar audibilidade de fala em um sinal de áudio de múltiplos canais, compreendendo:

5           comparar uma primeira característica e uma segunda característica do sinal de áudio de múltiplos canais para gerar um fator de atenuação, em que a primeira característica corresponde a um primeiro canal do sinal de áudio de múltiplos canais que contém áudio de fala e áudio não de fala, em que a primeira característica corresponde a uma primeira medida que é relacionada com uma intensidade de um sinal no primeiro canal, em que a  
10           segunda característica corresponde a um segundo canal do sinal de áudio de múltiplos canais que contém predominantemente áudio não de fala, e em que a segunda característica corresponde a uma segunda medida que é relacionada com uma intensidade de um sinal no segundo canal, incluindo:

15           determinar uma diferença entre a primeira medida e a segunda medida, e

            calcular o fator de atenuação com base na diferença entre a primeira medida e a segunda medida e em um valor limiar;

            ajustar o fator de atenuação de acordo com um valor de probabilidade de fala para gerar um fator de atenuação ajustado; e

20           atenuar o segundo canal usando o fator de atenuação ajustado.

2. Método de acordo com a reivindicação 1, compreendendo adicionalmente:

            processar o sinal de áudio de múltiplos canais para gerar a primeira característica e a segunda característica.

25           3. Método de acordo com a reivindicação 1, compreendendo adicionalmente:

            processar o primeiro canal para gerar o valor de probabilidade de fala.

30           4. Método de acordo com a reivindicação 1, em que o segundo canal é um de uma pluralidade de segundos canais, em que a segunda característica é uma de uma pluralidade de segundas características, em que o fator de atenuação é um de uma pluralidade de fatores de atenuação, e em

que o fator de atenuação ajustado é um de uma pluralidade de fatores de atenuação ajustados, compreendendo adicionalmente:

comparar a primeira característica e a pluralidade de segundas características para gerar a pluralidade de fatores de atenuação;

5                   ajustar a pluralidade de fatores de atenuação de acordo com o valor de probabilidade de fala para gerar a pluralidade de fatores de atenuação ajustados; e

atenuar a pluralidade de segundos canais usando a pluralidade de fatores de atenuação ajustados.

10                  5. Método de acordo com a reivindicação 1, em que o sinal de áudio de múltiplos canais inclui um terceiro canal que contém predominantemente áudio não de fala, compreendendo adicionalmente:

comparar a primeira característica e uma terceira característica para gerar um fator de atenuação adicional, em que a terceira característica  
15                   corresponde ao terceiro canal;

ajustar o fator de atenuação adicional de acordo com o valor de probabilidade de fala para gerar um fator de atenuação adicional ajustado; e

atenuar o terceiro canal usando o fator de atenuação ajustado.

20                  6. Método de acordo com a reivindicação 1, em que a primeira medida é um primeiro nível de potência do sinal no primeiro canal, em que a segunda medida é um segundo nível de potência do sinal no segundo canal, e em que a diferença é uma diferença entre o primeiro nível de potência e o segundo nível de potência.

25                  7. Método de acordo com qualquer uma das reivindicações 1 a 5, em que a primeira medida é uma primeira potência do sinal no primeiro canal, em que a segunda medida é uma segunda potência do sinal no segundo canal, e em que a diferença é uma razão entre a primeira potência e a segunda potência.

30                  8. Aparelho incluindo um circuito para aperfeiçoar audibilidade de fala em um sinal de áudio de múltiplos canais, compreendendo:

um circuito de comparação que é configurado para comparar uma primeira característica e uma segunda característica do sinal de áudio

de múltiplos canais para gerar um fator de atenuação, em que a primeira característica corresponde a um primeiro canal do sinal de áudio de múltiplos canais que contém áudio de fala e áudio não de fala, em que a primeira característica corresponde a uma primeira medida que é relacionada com uma intensidade de um sinal no primeiro canal, em que a segunda característica corresponde a um segundo canal do sinal de áudio de múltiplos canais que contém predominantemente o áudio não de fala, e em que a segunda característica corresponde a uma segunda medida que é relacionada com uma intensidade de um sinal no segundo canal, em que o circuito de comparação é configurado:

para determinar uma diferença entre a primeira medida e a segunda medida, e

para calcular o fator de atenuação com base na diferença entre a primeira medida e a segunda medida e em um valor limiar;

um multiplicador que é configurado para ajustar o fator de atenuação de acordo com um valor de probabilidade de fala para gerar um fator de atenuação ajustado; e

um amplificador que é configurado para atenuar o segundo canal usando o fator de atenuação ajustado.

9. Aparelho de acordo com a reivindicação 8, em que a primeira característica corresponde a um primeiro nível de potência e em que a segunda característica corresponde a um segundo nível de potência, e em que o circuito de comparação compreende:

um primeiro somador que é configurado para subtrair o primeiro nível de potência do segundo nível de potência para gerar uma diferença de nível de potência;

um segundo somador que é configurado para somar a diferença de nível de potência e o valor limiar para gerar uma margem; e

um circuito limitador que é configurado para calcular o fator de atenuação como um maior que um de a margem e zero.

10. Aparelho de acordo com a reivindicação 8, em que a primeira característica corresponde a um primeiro nível de potência e em que a

segunda característica corresponde a um segundo nível de potência, compreendendo adicionalmente:

um primeiro estimador de potência que é configurado para calcular o primeiro nível de potência do primeiro canal; e

5 um segundo estimador de potência que é configurado para calcular o segundo nível de potência do segundo canal.

11. Aparelho de acordo com a reivindicação 8, compreendendo adicionalmente:

10 um processador de determinação de fala que é configurado para processar o primeiro canal para gerar o valor de probabilidade de fala.

12. Programa de computador incorporado em mídia de gravação tangível para aperfeiçoar audibilidade de fala em um sinal de áudio de múltiplos canais, o programa de computador controlando um dispositivo para executar processamento compreendendo:

15 comparar uma primeira característica e uma segunda característica do sinal de áudio de múltiplos canais para gerar um fator de atenuação, em que a primeira característica corresponde a um primeiro canal do sinal de áudio de múltiplos canais que contém áudio de fala e áudio não de fala, em que a primeira característica corresponde a uma primeira medida que é

20 relacionada com uma intensidade de um sinal no primeiro canal, em que a segunda característica corresponde a um segundo canal do sinal de áudio de múltiplos canais que contém predominantemente o áudio não de fala, e em que a segunda característica corresponde a uma segunda medida que é relacionada com uma intensidade de um sinal no segundo canal, incluindo:

25 determinar uma diferença entre a primeira medida e a segunda medida, e

calcular o fator de atenuação com base na diferença entre a primeira medida e a segunda medida e em um valor limiar;

30 ajustar o fator de atenuação de acordo com um valor de probabilidade de fala para gerar um fator de atenuação ajustado; e

atenuar o segundo canal usando o fator de atenuação ajustado.

13. Aparelho para aperfeiçoar audibilidade de fala em um sinal

de áudio de múltiplos canais, compreendendo:

dispositivo para comparar uma primeira característica e uma segunda característica do sinal de áudio de múltiplos canais para gerar um fator de atenuação, em que a primeira característica corresponde a um primeiro canal do sinal de áudio de múltiplos canais que contém áudio de fala e áudio não de fala, em que a primeira característica corresponde a uma primeira medida que é relacionada com uma intensidade de um sinal no primeiro canal, em que a segunda característica corresponde a um segundo canal do sinal de áudio de múltiplos canais que contém predominantemente o áudio não de fala, e em que a segunda característica corresponde a uma segunda medida que é relacionada com uma intensidade de um sinal no segundo canal, incluindo:

dispositivo para determinar uma diferença entre a primeira medida e a segunda medida, e

dispositivo para calcular o fator de atenuação com base na diferença entre a primeira medida e a segunda medida e em um valor limiar;

dispositivo para ajustar o fator de atenuação de acordo com um valor de probabilidade de fala para gerar um fator de atenuação ajustado; e

dispositivo para atenuar o segundo canal usando o fator de atenuação ajustado.

14. Aparelho de acordo com a reivindicação 13, em que a primeira característica corresponde a um primeiro nível de potência e em que a segunda característica corresponde a um segundo nível de potência, em que o dispositivo para comparar compreende:

dispositivo para subtrair o primeiro nível de potência do segundo nível de potência para gerar uma diferença de nível de potência.

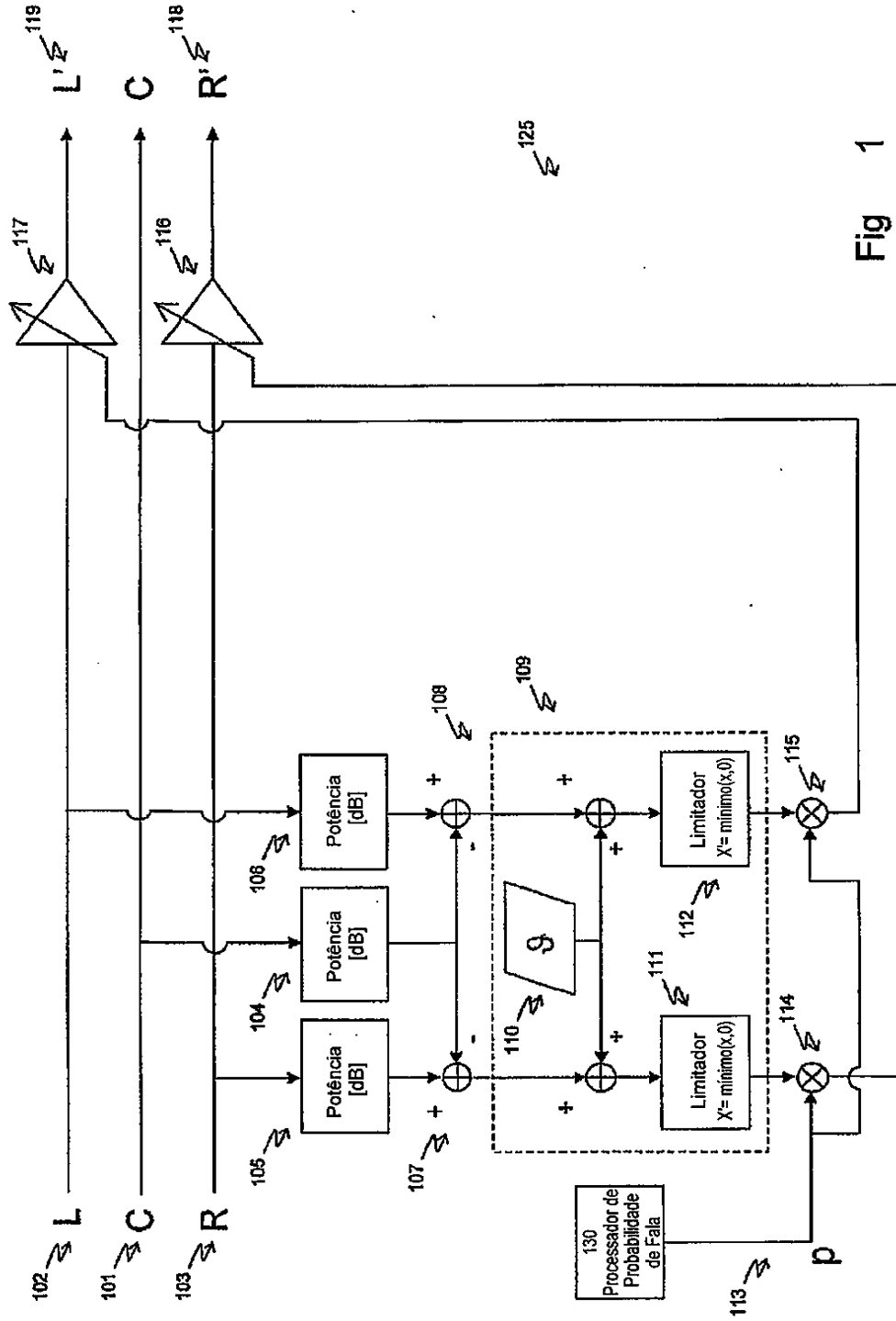


Fig 1

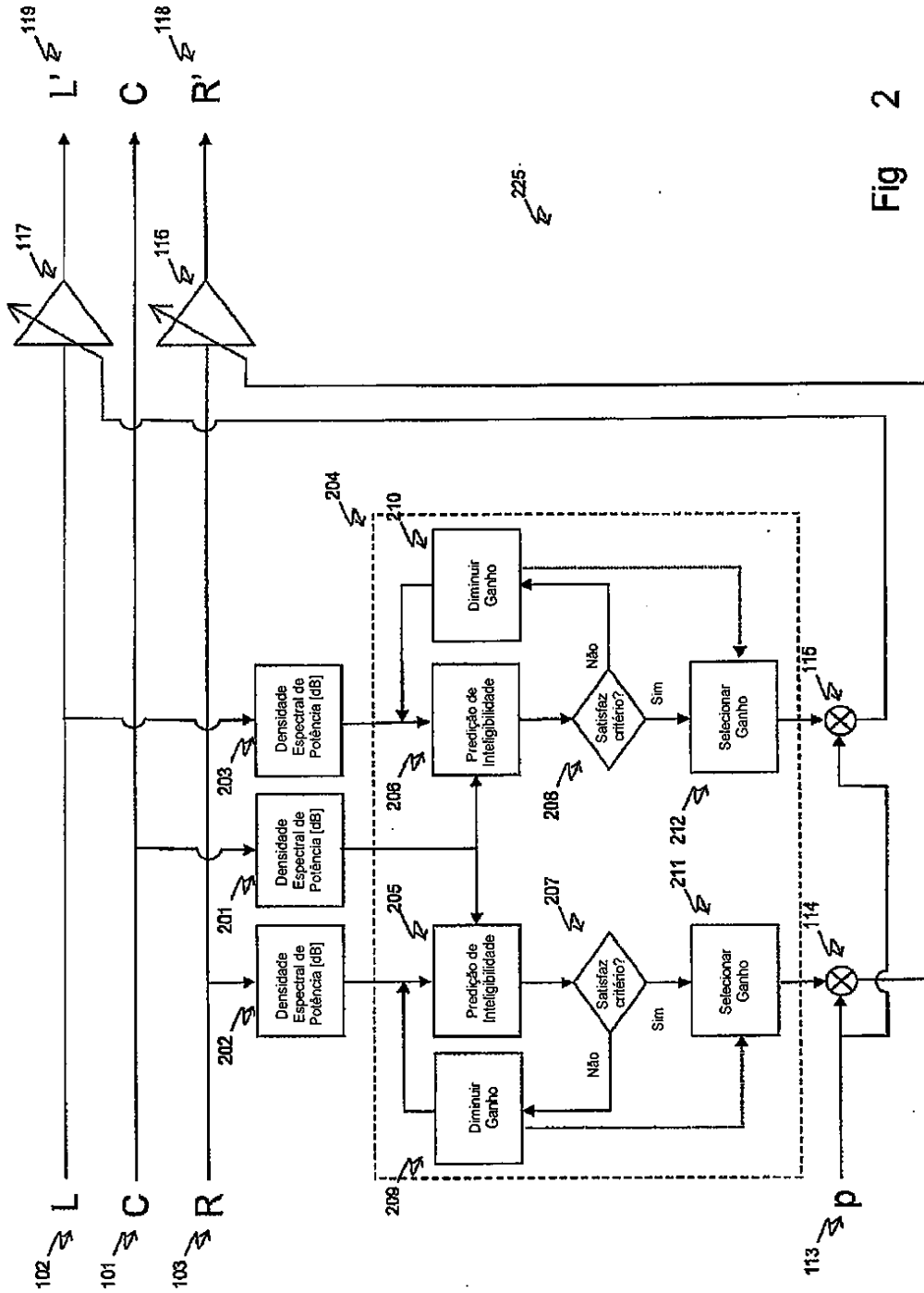


Fig 2

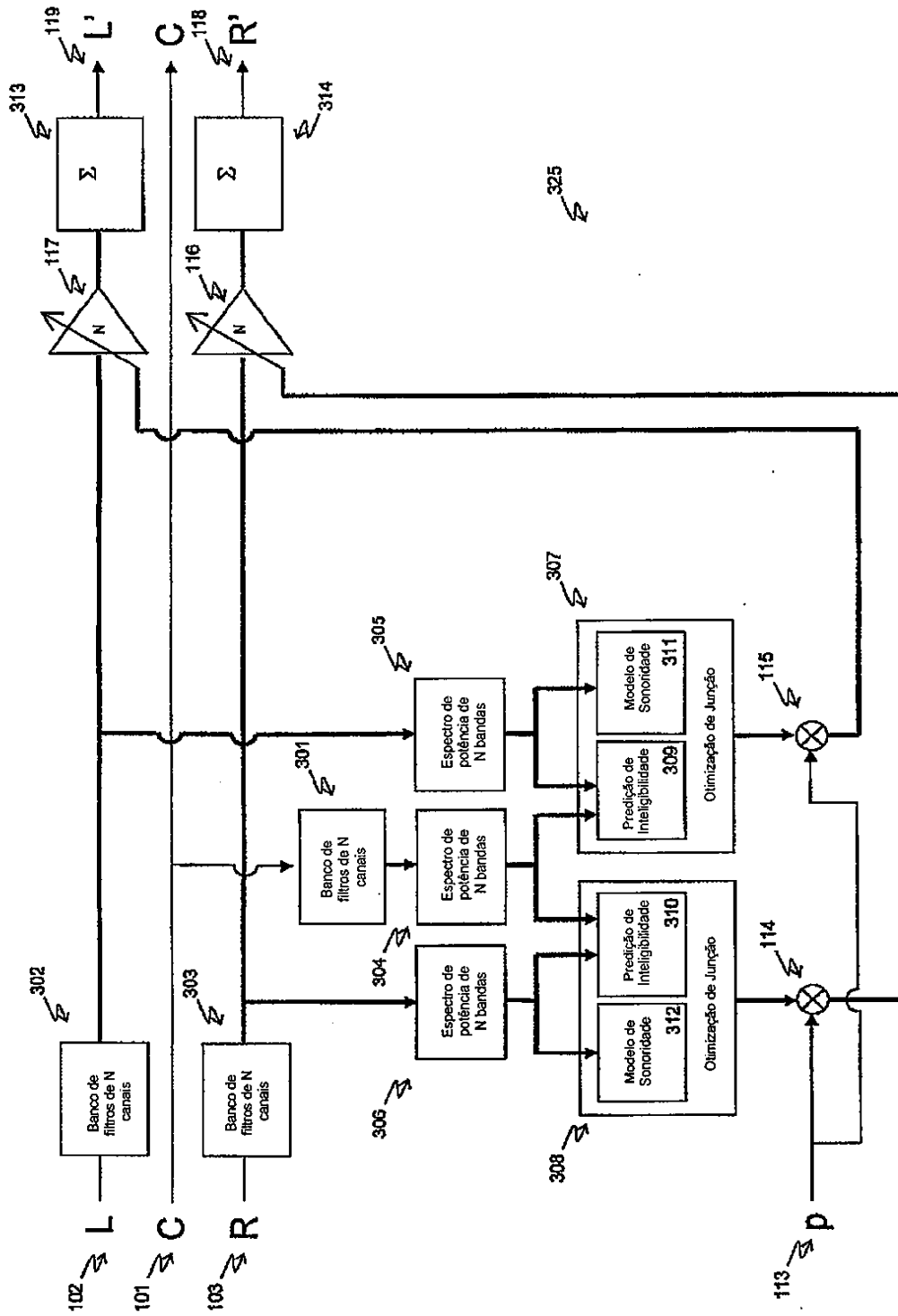


Fig 3

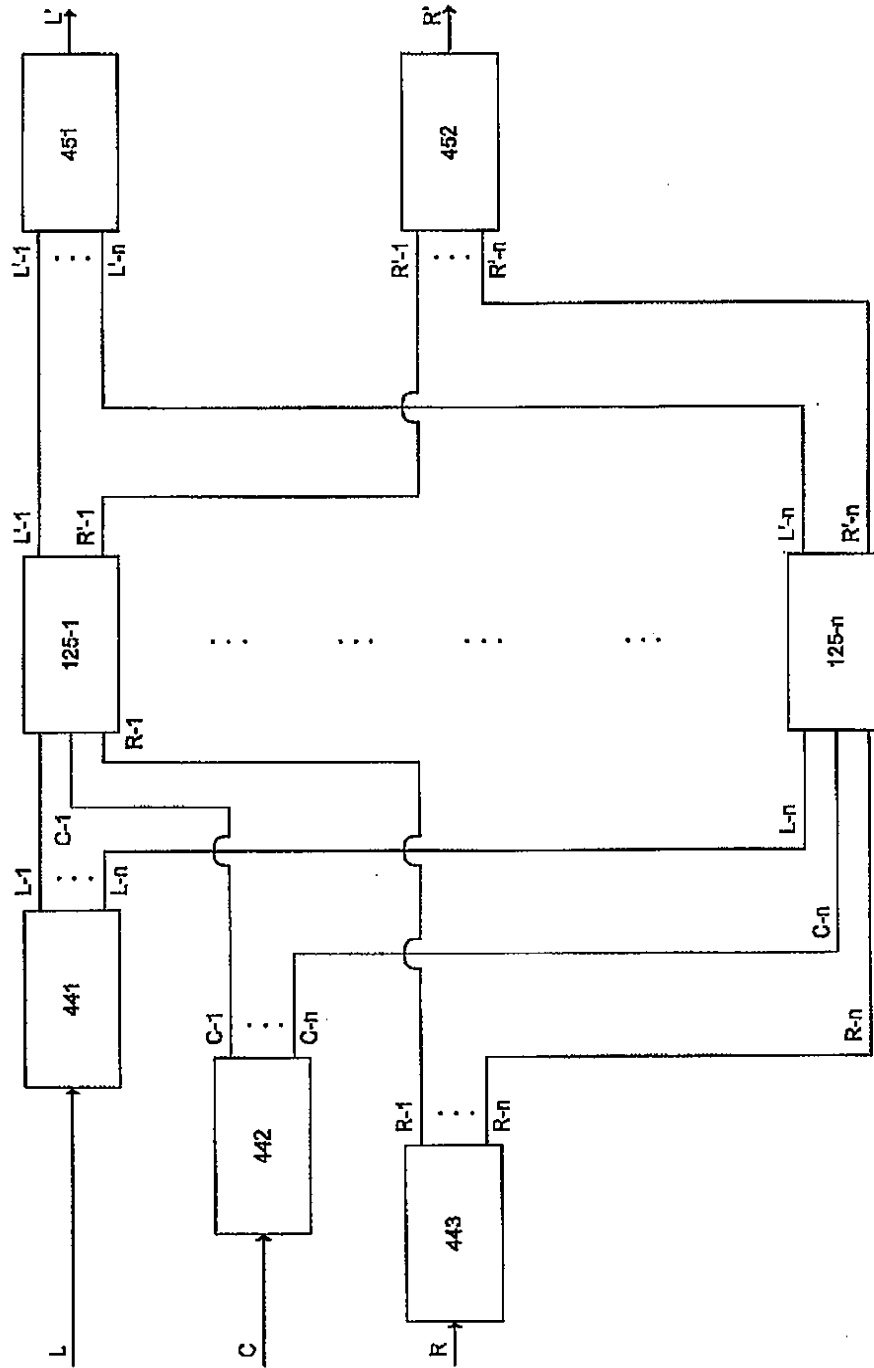


Fig 4A

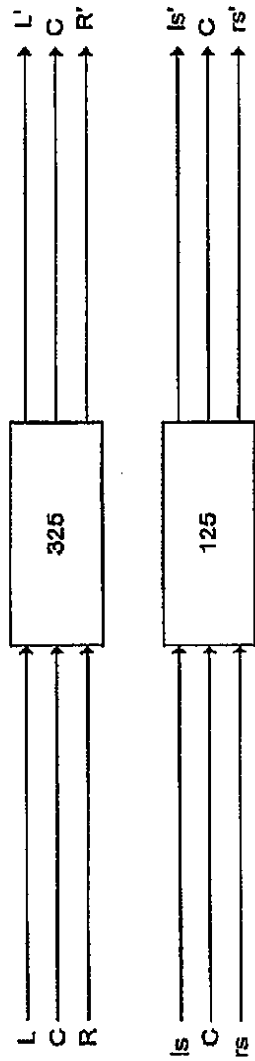


Fig 4B

## RESUMO

Patente de Invenção: **"MÉTODO E APARELHO PARA MANTER AUDIBILIDADE DE FALA EM ÁUDIO DE MÚLTIPLOS CANAIS COM IMPACTO MÍNIMO EM EXPERIÊNCIA ENVOLVENTE"**.

5           A presente invenção refere-se a um método de aperfeiçoar audibilidade de fala em um sinal de áudio de múltiplos canais. O método inclui comparar uma primeira característica e uma segunda característica do sinal de áudio de múltiplos canais para gerar um fator de atenuação. A primeira característica corresponde a um primeiro canal do sinal de áudio de múltiplos canais que contém áudio de fala e não fala, e a segunda característica  
10           corresponde a um segundo canal do sinal de áudio de múltiplos canais que contém predominantemente áudio de não fala. O método inclui adicionalmente ajustar o fator de atenuação de acordo com um valor de probabilidade de fala para gerar um fator de atenuação ajustado. O método inclui adicionalmente atenuar o segundo canal usando o fator de atenuação ajustado.  
15