



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

(11) BR 112016013366-8 B1



(22) Data do Depósito: 02/12/2014

(45) Data de Concessão: 28/12/2021

(54) Título: EXTRAÇÃO DE SOM REVERBERANTE UTILIZANDO REDES DE MICROFONES

(51) Int.Cl.: G10L 21/02; G10L 19/26; H04R 5/027; G11B 20/10.

(30) Prioridade Unionista: 11/12/2013 EP 13196672.3; 20/02/2014 EP 14156014.4.

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

(72) Inventor(es): OLIVER THIERGART; EMANUEL HABETS.

(86) Pedido PCT: PCT EP2014076252 de 02/12/2014

(87) Publicação PCT: WO 2015/086377 de 18/06/2015

(85) Data do Início da Fase Nacional: 09/06/2016

(57) Resumo: EXTRAÇÃO DE SOM REVERBERANTE UTILIZANDO REDES DE MICROFONES. Método compreende a estimativa de uma coerência espacial entre uma primeira parte de som difuso em um primeiro sinal de microfone e uma segunda parte de som difuso em um segundo sinal de microfone. O primeiro sinal de microfone é capturado por um primeiro microfone e o segundo sinal de microfone é capturado por um segundo microfone que se encontra afastado do primeiro microfone de uma maneira conhecida. O método compreende ainda a definição de uma restrição linear para coeficientes de filtro de um filtro de som difuso, sendo que a restrição linear é baseada na coerência espacial. O método compreende também o cálculo de pelo menos uma das estatísticas de sinal e estatísticas de ruído sobre o primeiro sinal de microfone e o segundo sinal de microfone. O método compreende também a determinação dos coeficientes de filtro do filtro de som difuso solucionando um problema de otimização relativo pelo menos a uma das estatísticas de sinal e estatísticas de ruído enquanto considera a restrição linear para os coeficientes de filtro.

Relatório Descritivo da Patente de Invenção para: **“EXTRAÇÃO DE SOM REVERBERANTE UTILIZANDO REDES DE MICROFONES”**.

[001] A invenção seguinte encontra-se na área da análise acústica, gravação de som espacial, processamento de sinais de redes de microfones, e filtragem espacial. Alguns modelos desta invenção dizem respeito a um método que pode ser utilizado para determinar os coeficientes do filtro de um filtro de som difuso, isto é, um filtro para extrair som difuso (som reverberante) das gravações com uma rede de microfones. Alguns modelos dizem respeito a um programa de computador correspondente. Alguns modelos dizem respeito a um aparelho que pode ser utilizado para determinar os coeficientes do filtro de um filtro de som difuso.

[002] A aquisição de som com redes de microfones em ambientes reverberantes tem habitualmente o objetivo de capturar o som direto das fontes de som enquanto atenua ruído e reverberação. Para muitas aplicações seria benéfico se existisse a possibilidade de extrair também o som reverberante enquanto se suprime o som e ruído direto. Por exemplo, na reprodução de som espacial [Pulkki2007, Thiergart2013, Kowalczyk2013], a reverberação presente no lado de gravação precisa ser reproduzida no lado de reprodução para criar a impressão espacial desejada. Além disso, dada uma estimativa do som reverberante, podemos calcular parâmetros tais como a taxa de sinal para a reverberação ou potência do som reverberante, que apresenta informação crucial para várias outras aplicações.

[003] Enquanto o cálculo de componentes de som direto (por ex.: utilizando separação na origem, de reverberação, ou redução de ruído) é bem tratado na literatura, existem apenas algumas abordagens para extração de som reverberante. Habitualmente, a reverberação é modelada como uma área de som difuso (variável no tempo). Para extrair o som difuso, filtros monocal foram recentemente utilizados (por ex.: em [Pulkki2007, Thiergart2013]), rendendo um fraco desempenho quando múltiplas fontes se encontram ativas ou para sinais do tipo transitório. Um melhor desempenho pode ser obtido com

filtros multicanal (por ex.: [Kowalczyk2013, Thiergart2013b]). Infelizmente, os filtros multicanais atualmente existentes não são ótimos e não rendem um padrão de diretividade adequado para capturar som difuso.

[004] Seria, pois, desejável fornecer um filtro de som difuso dotado de desempenho melhorado em termos de extração de som difuso e/ou supressão de som direto. Poderia ser também desejável que o filtro de som difuso fosse dotado de uma resposta direcional altamente omnidirecional, à exceção de direções de chegada de componentes de som direto. Uma resposta direcional altamente omnidirecional é desejada visto que o som difuso chega de todas as direções à rede de microfones.

[005] Um método é fornecido compreendendo a definição de uma restrição linear para coeficientes do filtro de um filtro de som difuso. A restrição linear é baseada em uma coerência espacial entre uma primeira parte de som difuso em um primeiro sinal do microfone e uma segunda parte de som difuso em um segundo sinal do microfone. O primeiro sinal do microfone é capturado por um primeiro microfone e o segundo sinal do microfone é capturado por um segundo microfone distanciado do primeiro microfone em uma maneira conhecida. O método compreende também o cálculo de pelo menos um de uma direção de chegada de pelo menos um som direto, estatística de sinal sobre o primeiro e o segundo sinais do microfone, e estatística de ruído sobre o primeiro e o segundo sinais do microfone. O método compreende ainda a determinação dos coeficientes do filtro do filtro de som difuso solucionando um problema de otimização relativamente a pelo menos uma direção de chegada de pelo menos um som direto, a estatística de sinal, e a estatística de ruído enquanto considerando a restrição linear para os coeficientes de filtro.

[006] Modelos fornecem um programa de computador para implementação do método descrito em cima quando é executado em um computador ou o processador de sinal é fornecido.

[007] Modelos adicionais fornecem um aparelho compreendendo um calculador de restrição linear configurado para definir uma restrição linear para

coeficientes de filtro de um filtro de som difuso. A restrição linear é baseada em uma coerência espacial entre uma primeira parte de som difuso em um primeiro sinal do microfone e uma segunda parte de som difuso em um segundo sinal do microfone. O primeiro sinal do microfone é ou tem sido capturado por um primeiro microfone e o segundo sinal do microfone é ou tem sido capturado por um segundo microfone afastado do primeiro microfone de uma maneira conhecida. O aparelho compreende também um calculador de estatística, configurado para calcular pelo menos um de uma direção de chegada de pelo menos um som direto, estatística de sinal sobre o primeiro e o segundo sinal do microfone, e estatísticas de ruído sobre o primeiro e o segundo sinal do microfone e o segundo sinal do microfone. O aparelho compreende ainda um calculador de coeficientes do filtro configurado para determinar os coeficientes de filtro do filtro de som difuso solucionando um problema de otimização relativamente a pelo menos um da direção de chegada do pelo menos um som direto, a estatística do sinal, e a estatística de ruído enquanto considera a restrição linear para os coeficientes de filtro.

[008] Modelos são baseados no discernimento que um filtro de som difuso pode ser determinado enquanto tem em consideração pelo menos uma restrição linear relativa às partes do som difuso dos sinais do microfone.

[009] A seguir, modelos desta invenção são descritos com maior detalhe com referência às figuras, nas quais:

[010] Fig. 1 ilustra um esquema de diagrama de blocos para extrair som difuso com um filtro monocanal;

[011] Fig. 2 ilustra um esquema de diagrama de blocos de uma abordagem para extrair som difuso com um filtro de monocanal;

[012] Fig. 3 ilustra um esquema de diagrama de blocos da invenção proposta de acordo com um primeiro exemplo para implementação;

[013] Fig. 4 ilustra um esquema de diagrama de blocos da invenção proposta de acordo com um segundo exemplo para implementação;

[014] Fig. 5 ilustra um esquema de diagrama de blocos da invenção

proposta de acordo com um terceiro exemplo para implementação;

[015] Fig. 6 ilustra um exemplo de um padrão transmissor resultante da abordagem para extração de som difuso com um filtro de acordo com a Fig. 2;

[016] Fig. 7 ilustra um exemplo de um padrão transmissor resultante da abordagem para extração de som difuso com um filtro de acordo com a Fig. 4; e

[017] Fig. 8 ilustra

[018] Os termos “som direto” e “som difuso” são definidos do seguinte modo.

[019] Som direto: sons que chegam a partir principalmente de uma direção visível específica nos microfones. O som direto pode representar por exemplo o som que viaja diretamente da fonte sonora para o microfone ou uma reflexão de espaço distinto. Sons diretos podem ser por exemplo ondas planas ou ondas esféricas com uma direção de chegada específica. Quando a direção de chegada de um som direto é conhecida, pode-se calcular a função de transferência relativa do som direto entre os microfones visto que a geometria do microfone é conhecida.

[020] Som difuso: som que chega aos microfones de todas as direções. O som difuso pode representar por exemplo a última reverberação presente em um espaço. Habitualmente, nenhuma direção de chegada visível pode ser associada a um som difuso (campo sonoro isotrópico), isto é, o som chega com potência média igual de todas as direções. Além disso, as funções de transferência relativa do som difuso entre os microfones devem ser assumidas aleatórias e não observáveis. Contudo, as funções de transferência média relativa do som difuso entre os microfones são habitualmente conhecidas para definições específicas de microfone e difundem suposições do campo ou podem ser medidas.

[021] As seguintes subseções resumem abordagens resumidas para extrair som difuso (ou som reverberante) de uma gravação de microfone. A seguir, M indica o número de microfones utilizados. Pressupõe-se que todos os

sinais de microfones foram transformados no domínio tempo-frequência em que k é o índice de frequência e n é o índice tempo (de salientar que os filtros podem habitualmente ser também aplicados no domínio tempo). Os microfones capturam L ondas planas (referidas como som direto) propagando em um campo difuso. O DOA do l^{a} onda plana é representado pelo vetor normal reduzido $\mathbf{n}_l(k, n)$. No domínio tempo-frequência, o sinal do m -th microfone (omnidirecional) pode ser escrito como

$$X_m(k, n) = \sum_{l=1}^L X_l(k, n, \mathbf{d}_m) + X_d(k, n, \mathbf{d}_m) + X_n(k, n, \mathbf{d}_m)$$

[022] Aqui, $X_l(k, n)$ é a pressão sonora da l^{a} onda plana, $X_d(k, n, \mathbf{d}_m)$ é o som difuso, $X_n(k, n, \mathbf{d}_m)$ é um ruído estacionário (por ex.: ruído próprio ou ruído de fundo), e $\mathbf{d}_m$ é um vetor descrevendo a posição do microfone (do m -th microfone) em um dado sistema de coordenadas.

[023] O objetivo desta invenção é estimar $X_d(k, n, \mathbf{d}_m)$ na posição $\mathbf{d}_m$.

[024] Filtros monocanal extraem o som difuso de um único sinal de microfone ($M = 1$). Tais filtros são utilizados por exemplo na Codificação Áudio Direcional [Pulkki2007] ou no Microfone Virtual [Thiergart2013].

[025] Uma estimativa do som difuso é encontrada multiplicando um dos sinais de microfone, por exemplo o sinal de microfone do primeiro microfone $X_1(k, n)$, com um filtro $H(k, n)$, por ex.:

$$\hat{X}_d(k, n, \mathbf{d}_m) = X_1(k, n)H(k, n)$$

[026] Habitualmente, o filtro $H(k, n)$ é um filtro Wiener, dado por

$$H(k, n) = \frac{\phi_d(k, n)}{\phi_d(k, n) + \phi_u(k, n)}$$

[027] em que ϕ_d é a potência do som difuso e ϕ_u é a potência das ondas planas e do ruído estacionário. Nalgumas aplicações, a raiz quadrada do filtro Wiener (ou seja, a raiz quadrada de H) é utilizada em vez do filtro Wiener. De salientar que para calcular $H(k, n)$, temos que calcular a potência ϕ_d e ϕ_u . Para

esta finalidade, podemos considerar por exemplo a taxa de sinal a difundir (SDR), que pode ser calculada como explicado em [Thiergart2012]. Em alternativa, $H(k, n)$ pode ser encontrado calculando a chamada redundância, tal como descrito em, [Pulkki2007,Thiergart2013]. Calculando a SDR ou redundância habitualmente exige mais de um microfone. Não obstante, o som difuso é finalmente obtido filtrando um único sinal de microfone.

[028] Um sistema exemplo para extrair o som difuso com um filtro monocanal é ilustrado na Fig. 1. Em primeiro lugar, a SDR (ou alternativamente a redundância) é calculada a partir de sinais de múltiplos microfones. Em segundo lugar, o filtro $H(k, n)$ é calculado a partir desta informação. Finalmente, o filtro $H(k, n)$ é multiplicado com o sinal de um único microfone para obter o cálculo do som difuso.

[029] Filtros multicanal consideram microfones $M > 1$. Tais filtros têm sido utilizados por exemplo em [Thiergart2013b, Kowalczyk2013]. Para as seguintes derivações, deixem-nos representar os sinais de microfone M através de um vetor $\mathbf{x}(k, n) = [X_1(k, n), X_2(k, n), \dots, X_M(k, n)]^T$. O som difuso no m -th microfone é calculado através de uma combinação linear dos sinais de microfone M , ou seja,

$$\hat{X}_d(k, n, \mathbf{d}_m) = \mathbf{w}_m^H(k, n) \mathbf{x}(k, n)$$

[030] em que $\mathbf{w}_m$ é um vetor de peso complexo de comprimento M . As ponderações $\mathbf{w}_m$ deverão ser calculadas de modo que um cálculo estimado do som difuso seja obtido.

[031] A maneira mais direta de descobrir um filtro adequado é calcular as ponderações $\mathbf{w}_m$ de modo que as ondas planas L sejam suprimidas enquanto o ruído estacionário $X_n(k, n, \mathbf{d}_m)$, contido nos sinais de microfone, é minimizado. Matematicamente expresso, as ponderações do filtro são dadas como

$$\mathbf{w}_m(k, n) = \underset{\mathbf{w}}{\operatorname{argmin}} \quad \mathbf{w}^H \Phi_n(k) \mathbf{w}$$

[032] sujeito às restrições lineares

$$\mathbf{w}^H \mathbf{a}_l(k, n) = 0 \quad \forall l$$

[033] Aqui, Φ_n é a matriz PSD (matriz de densidade espectral de potência) do ruído estacionário, ou seja, $\Phi_n = E\{\mathbf{x}_n \mathbf{x}_n^H\}$, que pode ser calculada com abordagens bem conhecidas por exemplo quando nenhum som difuso ou som direto se encontra presente. Além disso, $\mathbf{a}_l$ é o chamado vetor de propagação. Os seus elementos são a função de transferência relativa da l^a onda plana do m -th microfone para os outros microfones. Assim, $\mathbf{a}_l$ é um vetor coluna com comprimento M (recordata: apenas o som difuso no m -th microfone é calculado pela combinação linear ponderada- $\mathbf{w}_m$ dos sinais de microfone M ; o som difuso nos outros microfones é significativamente redundante, pois estes sinais estão relacionados através das funções de transferência relativa do m -th microfone para os outros microfones e pode ser calculado desta maneira, caso necessário). Os elementos de $\mathbf{a}_l$ dependem do DOA da l^a onda plana. Isto significa que $\mathbf{a}_l$ é uma função do DOA da l^a onda plana, ou seja, $\mathbf{a}_l = f(\mathbf{n}_l)$. Visto que $\mathbf{a}_l$ depende do som direto (ou seja, ondas planas), é preferido como restrição do som direto como segue. Com este filtro especial criamos essencialmente um gerador de feixes, dotado de um padrão transmissor com nulos nas direções das direções das ondas planas L . Em resultado disso, todas as ondas planas são suprimidas. Infelizmente, solucionando este problema de minimização em cima conduz a zero ponderações $\mathbf{w}_m$ visto que temos apenas zero restrições, ou seja, o som difuso não pode ser extraído.

[034] Para ultrapassar este problema e evitar zero ponderações de filtro, [Thiergart2013b, Kowalczyk2013] propõe utilizar o mesmo filtro, mas com uma restrição adicional, dada por

$$\mathbf{w}^H \mathbf{a}_0(k, n) = 1$$

[035] em que $\mathbf{a}_0$ é um vetor de propagação que corresponde a um DOA específico $\mathbf{n}_0$, a partir do qual nenhuma onda plana chega. Com esta restrição

evita-se zero ponderações de filtro, mas ainda não se captura o indesejado som direto. Em resultado disso, com este filtro apenas o som difuso e algum ruído é capturado, mas todas as ondas planas são atenuadas. Em [Thiergart2013b], o DOA $\mathbf{n}_0$, ao qual corresponde o vetor $\mathbf{a}_0$, escolhe a direção que tem a maior distância angular a todos os DOAs $\mathbf{n}_l(k, n)$ das ondas planas. Por exemplo, se uma única onda plana chega de 0 graus, então $\mathbf{n}_l(k, n)$ iria corresponder a 180 graus. Infelizmente, o DOA $\mathbf{n}_0$ não garante que obtemos um cálculo do som difuso com o menos ruído possível. Além disso, o padrão transmissor resultante não é muito ótimo para capturar som difuso, visto que se torna altamente diretivo nas frequências mais elevadas. Isto é um inconveniente quando o objetivo é capturar som difuso de todas as direções.

[036] Um exemplo de um padrão transmissor resultante é indicado na Fig. 6. Aqui, dois sons diretos chegam de uma direção azimutal de 51° e 97° . A figura ilustra o padrão transmissor resultante em uma frequência de 2,8kHz quando utiliza uma rede linear uniforme com 16 microfones com espaçamento de microfone de 5 cm. O padrão transmissor é dotado de nulos exatamente para 51° e 97° e um ganho elevado para 180° , correspondendo à direção $\mathbf{n}_0$. Além disso, o padrão transmissor possui múltiplos outros nulos espaciais ou baixos ganhos para quase todas as outras direções. Este padrão transmissor não é adequado para capturar som difuso que chega de todas as direções. De salientar novamente que as restrições do som direto $\mathbf{a}_l$ estão diretamente relacionadas com os DOAs dos sons diretos. Um padrão transmissor desejado, que não pode ser obtido com o filtro espacial nesta subseção, encontra-se indicado na Fig. 7. Este padrão transmissor possui dois nulos espaciais para os DOAs dos sons diretos, mas caso contrário é quase omnidirecional. Este padrão transmissor é obtido utilizando-se o filtro proposto descrito em baixo relativamente à Fig. 7.

[037] Uma solução em forma fechada para calcular as ponderações do filtro $\mathbf{w}_m$ dadas nas restrições em cima pode ser encontrada em

[VanTrees2002]. De modo a calcular o filtro especial, deveremos conhecer o DOA das ondas planas L , em especial para calcular as restrições do som direto $\mathbf{a}_l$ e $\mathbf{a}_0$. Esta informação DOA pode ser determinada com estimadores DOA de banda estreita bem conhecidos, tais como Root MUSIC ou ESPRIT. De salientar também que os elementos de $\mathbf{a}_0$ são habitualmente complexos e $\mathbf{a}_0$ habitualmente precisa ser calculado para cada k e n , visto que os DOAs das ondas planas deverão ser assumidos como altamente variáveis no tempo. O $\mathbf{a}_0$ altamente flutuante pode conduzir a perturbações sonoras.

[038] Um sistema exemplo para extrair o som difuso com o filtro multicanal apresentado é ilustrado na Fig. 2. Após transformar os sinais de microfone no domínio de tempo-frequência, calculamos o ruído estacionário e o DOA das ondas planas L . A partir da informação DOA as restrições lineares de som direto $M+1$ ($\mathbf{a}_l$ e $\mathbf{a}_0$) são então obtidas. Baseado nesta informação, as ponderações do filtro podem ser calculadas. A aplicação destas ponderações aos sinais do microfone rende o cálculo desejado do som difuso. Está claro a partir desta descrição que o filtro obtido depende apenas do som direto (ou seja, nos DOAs e nas respectivas funções de transferência relativa das ondas planas entre os microfones, respetivamente), mas não no som difuso. Isto significa que o filtro não considera informação potencialmente disponível no som difuso, mesmo que seja utilizada para calcular o som difuso.

[039] Nesta invenção propomos um filtro multicanal inovador para extrair som reverberante que ultrapasse as limitações dos filtros anteriormente referidos. O filtro espacial proposto é caracterizado por um padrão de diretividade, que tende para um padrão omnidirecional, exceto para as direções de chegada (DOAs) do som direto para o qual exhibe nulos espaciais. Isto representa uma propriedade altamente desejada para capturar som difuso de todas as direções com baixa distorção.

[040] A seguir propomos um filtro multicanal para calcular o som difuso $X_d(k, n, \mathbf{d}_m)$ na posição $\mathbf{d}_m$ com $M > 1$ microfones. No que diz respeito aos

filtros multicanal descritos em cima, a pressão do som difuso no m-th microfone é calculado desempenhando uma combinação linear dos sinais dos microfones, ou seja,

$$\hat{X}_d(k, n, \mathbf{d}_m) = \mathbf{w}_m^H(k, n) \mathbf{x}(k, n).$$

[041] O vetor de ponderação $\mathbf{w}_m$, a seguir proposto, minimiza uma função de custo específica e é linearmente restrita de modo idêntico aos filtros multicanal descritos em cima.

[042] Contudo, ao contrário dos filtros multicanal descritos em cima, propomos utilizar uma restrição linear que não depende do som direto (ou seja, nas ondas planas L). Mais precisamente, a restrição inovadora proposta não é uma função dos DOAs das ondas planas ou as funções de transferência relativa correspondente das ondas planas entre os microfones, respetivamente.

[043] Pelo contrário, a restrição inovadora proposta depende da informação estatística do som difuso, ou seja, a restrição inovadora proposta depende das funções de transferência relativa do som difuso entre os microfones. Segue-se uma ilustração em como a restrição inovadora proposta é uma função da coerência ou correlação do som difuso entre os microfones. Esta coerência corresponde a uma função de transferência média relativa do som difuso entre os microfones.

[044] O filtro espacial proposto é obtido minimizando uma função de custo específica enquanto satisfaz uma restrição sem distorção para o som difuso. Esta restrição corresponde à função de transferência relativa do som difuso entre os microfones. Matematicamente expresso, o filtro é calculado como

$$\mathbf{w}_m(k, n) = \underset{\mathbf{w}}{\operatorname{argmin}} J(\mathbf{w})$$

[045] sujeito à restrição linear

$$\mathbf{w}^H \mathbf{b}_m(k, n) = 1.$$

[046] Aqui, J é a função de custo a ser minimizada pelo filtro. A função de custo pode ser por exemplo a potência do ruído estacionário no filtro de saída, a energia interferente no filtro de saída, ou o erro quadrático do som difuso

estimado. Exemplos para J serão fornecidos nos modelos. O vetor de restrição $\mathbf{b}_m$ é dado por $\mathbf{b}_m(k, n) = [B_{1,m}(k, n), B_{2,m}(k, n), \dots, B_{M,m}(k, n)]^T$. O m -th elemento $B_{m',m}$ é a função de transferência relativa do som difuso entre microfone m e m' . Esta função de transferência relativa é dada por

$$B_{m',m}(k, n) = \frac{X_d(k, n, \mathbf{d}_m)}{X_d(k, n, \mathbf{d}_{m'})}.$$

[047] De salientar que o m -th elemento de $\mathbf{b}_m$ é igual a 1. Com esta restrição capturamos o som difuso ser distorção. Na verdade, que $\mathbf{x}_d(k, n) = [X_d(k, n, \mathbf{d}_1), X_d(k, n, \mathbf{d}_2), \dots, X_d(k, n, \mathbf{d}_M)]^T$ seja um vetor contendo o som difuso gravado. Com as equações em cima, este vetor pode ser redigido como

$$\mathbf{x}_d(k, n) = \mathbf{b}_m(k, n) X_d(k, n, \mathbf{d}_m).$$

[048] O som difuso à saída do filtro é dado por $\mathbf{w}^H(k, n) \mathbf{x}_d(k, n)$, idêntico a $X_d(k, n, \mathbf{d}_m)$ visto que $\mathbf{w}^H \mathbf{b}_m(k, n) = 1$. Desse modo, este filtro captura o som difuso sem distorção. As funções de transferência relativa em $\mathbf{b}_m$ não podem habitualmente ser estimadas na prática pois são basicamente aleatórias, ou seja, temos uma percepção diferente da função de transferência para cada k e n . Assim, na prática, $B_{m',m}$ é calculado como a função media relativa entre o microfone m e m' , ou seja,

$$B_{m',m}(k, n) = \gamma_{m',m}(k, n).$$

[049] Esta função de transferência media relativa $\gamma_{m',m}$ corresponde à chamada coerência especial do som difuso entre o microfone m e m' , definido como

$$\gamma_{m',m}(k, n) = \frac{E\{X_d(k, n, \mathbf{d}_m) X_d^*(k, n, \mathbf{d}_{m'})\}}{\sqrt{E\{|X_d(k, n, \mathbf{d}_m)|^2\} E\{|X_d(k, n, \mathbf{d}_{m'})|^2\}}}.$$

[050] em que $(\cdot)^*$ indica conjugado complexo. Esta coerência especial descreve a correlação especial do som difuso entre o microfone m e m' no domínio de frequência. Esta coerência depende do campo sonoro difuso específico. A coerência pode ser medida com antecedência para um dado espaço. Alternativamente, a coerência é conhecida da teoria para campos sonoros difusos específicos [Elko2001]. Por exemplo para um campo sonoro difuso esfericamente isotrópico, que pode frequentemente ser assumido na prática, temos

$$\gamma_{m',m}(k,n) = \text{sinc}\left(2\pi \frac{f}{c} r_{m',m}\right)$$

[051] em que sinc indica a função sinc, f é a frequência acústica para a dada banda de frequência k , e c é a velocidade do som. Além disso, $r_{m',m}$ é a distância entre microfones m e m' . Quando utilizando a coerência espacial como a restrição linear $B_{m',m}$, que representa a função de transferência média relativa do som difuso entre os microfones, então o filtro obtido é equivalente à soma de muitos filtros espaciais linearmente restritos, em que cada um destes filtros captura uma diferente realização do som difuso aleatório sem distorção.

[052] Com a restrição de som difuso introduzido em cima, obtemos um filtro espacial que captura o som difuso igualmente bem de todas as direções. Isto é ao contrário dos filtros multicanal descritos em cima, que capturam o som principalmente de uma direção, em especial da direção à qual corresponde o vetor de propagação escolhido $\mathbf{a}_0$.

[053] De salientar que a restrição de som difuso $\mathbf{b}_m$ é conceptualmente muito diferente das restrições de som direto $\mathbf{a}_l$ e $\mathbf{a}_0$. Desse modo, o filtro inovador proposto nesta seção é conceptualmente muito diferente comparado aos filtros multicanal descritos em cima.

[054] Um esquema de blocos da invenção proposta é indicado na Fig. 3. Primeiro, os sinais de microfone M são transformados no domínio de tempo-

frequência (ou outro domínio adequado para o processamento do sinal) utilizando um banco de filtros (FB) (101). Segundo, calculamos o vetor de restrição linear de som difuso $\mathbf{b}_m$ no bloco (102). O vetor de restrição linear de som difuso é estimado a partir do sinal ou corresponde por exemplo à coerência espacial teórica para um campo difuso específico assumido tal como mencionado antes. No bloco (104), estimamos estatística específica (por ex.: estatística de ruído) dos sinais de microfone. Esta informação, habitualmente representada como uma matriz PSD $\Phi(k, n)$, é utilizada para gerar a função de custo J que tem de ser minimizada pelo filtro. As ponderações do filtro que minimizam a função de custo sujeita à restrição de som difuso são calculadas em bloco (103). Finalmente, as ponderações são aplicadas aos sinais de microfone resultando na estimativa de som difuso desejado. Realizações específicas da invenção são apresentadas nos modelos seguintes.

[055] Minimização da Potência de Saída Satisfazendo uma Restrição de Som Difuso

[056] Neste modelo, define-se um filtro espacial que minimiza toda a saída do filtro sujeito à restrição de som difuso. A restrição de som difuso assegura que o som difuso é preservado pelo filtro espacial enquanto as restantes partes do sinal (ruído estacionário indesejado e ondas planas) são minimizadas. As ponderações do filtro $\mathbf{w}_m$ são calculadas como

$$\mathbf{w}_m(k, n) = \underset{\mathbf{w}}{\operatorname{argmin}} \mathbf{w}^H \Phi_x(k, n) \mathbf{w}$$

[057] sujeitas à restrição linear

$$\mathbf{w}^H \mathbf{b}_m(k, n) = 1$$

[058] Uma solução de forma fechada a este filtro é dada por [VanTrees2002]

$$\mathbf{w}_m^H(k, n) = \frac{\mathbf{b}_m^H(k, n) \Phi_x^{-1}(k, n)}{\mathbf{b}_m^H(k, n) \Phi_x^{-1}(k, n) \mathbf{b}_m(k, n)}$$

[059] Aqui, Φ_x é a matriz PSD dos sinais de microfone, que podem ser

calculados como

$$\Phi_x(k, n) = E\{\mathbf{x}(k, n) \mathbf{x}^H(k, n)\},$$

[060] em que $\mathbf{x}(k, n)$ é o vetor contendo os sinais de microfone. Na prática, a expectativa é aproximada por exemplo por uma média temporal. Além disso, os elementos do vetor de restrição $\mathbf{b}_m(k, n) = [B_{1,m}(k, n), B_{2,m}(k, n), \dots, B_{M,m}(k, n)]^T$ correspondem à coerência espacial do som difuso entre microfone m e m' , ou seja,

$$B_{m',m}(k, n) = \gamma_{m',m}(k, n).$$

[061] Na verdade, a coerência espacial $B_{m',m}$ não precisa depender do tempo (ou seja, $B_{m',m}(k, n) = B_{m',m}(k)$), visto que a coerência espacial pode ser estimada com antecedência ou assumir um valor teórico. A coerência espacial pode ser estimada a partir dos sinais de microfone (durante períodos em que apenas o som difuso está presente) utilizando

$$\gamma_{m',m}(k, n) = \frac{E\{X(k, n, \mathbf{d}_m) X^*(k, n, \mathbf{d}_{m'})\}}{\sqrt{E\{|X(k, n, \mathbf{d}_m)|^2\} E\{|X(k, n, \mathbf{d}_{m'})|^2\}}},$$

[062] ou considerado como uma informação prioritária assumindo um campo de som difuso específico. No último caso, podemos utilizar por exemplo a coerência espacial para um campo de som difuso esfericamente isotrópico, ou seja,

$$\gamma_{m',m}(k, n) = \text{sinc}\left(2\pi \frac{f}{c} r_{m',m}\right).$$

[063] De salientar que a função sinc pode ser substituída por outras funções dependendo do campo de som assumido. Para diferentes campos de som difuso existem diferentes funções de coerência conhecidas à priori. Exemplos podem ser encontrados em [Elko2001].

[064] Um esquema de blocos deste modelo é ilustrado na Fig. 3. Após transformação dos sinais de microfone com um banco de filtros (101),

calculamos a matriz do sinal PSD Φ_x no bloco de estimativa de estatística do sinal (104). Além disso, calculamos o vetor de restrição linear de som difuso $\mathbf{b}_m$ no bloco (102) a partir do sinal ou utilizando à priori informação assumindo uma campo de som difuso específico. As ponderações do filtro são então calculadas no bloco (103). Multiplicando as ponderações com os sinais de microfone rende a estimativa desejada do som difuso.

[065] O filtro calculado neste modelo possui as seguintes vantagens comparadas a outro filtro espacial (por ex.: os filtros descritos na área antecedente):

- Visto que minimizamos todo o sinal de saída, o filtro fornece um compromisso ótimo entre a atenuação das ondas planas L e o ruído estacionário.
- A matriz PSD do sinal exigido Φ_x , e logo o filtro, é mais fácil de calcular na prática.
- Nenhuma informação DOA das ondas planas L é exigida.

[066] Filtro de Variância Mínima Linearmente Restrito

[067] Este modelo representa uma combinação da abordagem inovadora e da abordagem do estado da arte de filtros multicanal descritos em cima relativamente à Fig. 2. Neste modelo, definimos um filtro espacial linearmente restrito que minimiza o ruído estacionário no filtro de saída sujeito à restrição difusa e restrições direcionais adicionais. As ponderações do filtro $\mathbf{w}_m$ são calculadas como

$$\mathbf{w}_m(k, n) = \underset{\mathbf{w}}{\operatorname{argmin}} \mathbf{w}^H \Phi_n(k) \mathbf{w}$$

[068] sujeitas às restrições lineares

$$\mathbf{w}^H \mathbf{b}_m(k, n) = 1$$

[069] e

$$\mathbf{w}^H \mathbf{a}_l(k, n) = 0 \quad \forall l$$

[070] Claramente, o filtro minimiza apenas o ruído estacionário na saída.

As ondas planas indesejadas são suprimidas com as segundas restrições lineares (tal como explicado em cima para os filtros multicanal, Fig. 2). Estas restrições adicionais comparadas com o filtro de minimização da potência de saída de acordo com a Fig. 3 assegura uma supressão ainda mais forte das ondas interferentes planas. O filtro resultante preserva ainda o som difuso devido à primeira restrição linear. Uma solução de forma fechada para este filtro, que pode ser calculada na prática, é dada por

$$\mathbf{w}_m^H(k, n) = \mathbf{g}^H(k, n) \left[\mathbf{C}^H(k, n) \Phi_n^{-1}(k) \mathbf{C}(k, n) \right]^{-1} \mathbf{C}^H(k, n) \Phi_n^{-1}(k)$$

[071] Aqui, o vetor $\mathbf{C} = [\mathbf{b}_m, \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_L]$ é a matriz de restrição contendo as restrições lineares em cima e $\mathbf{g} = [1, \mathbf{0}]^T$ ($\mathbf{0}$ um vetor de comprimento zero L) são as respectivas respostas. No que diz respeito ao filtro multicanal ilustrado na Fig. 2, os vetores $\mathbf{a}_l$ dependem dos DOAs das ondas planas L e podem ser calculados como bem conhecido pela literatura [VanTrees2002]. Pelo contrário, os elementos de $\mathbf{b}_m$ descrevem a correlação ou coerência do som difuso entre os microfones. Os elementos de $\mathbf{b}_m$ são calculados tal como explicado relativamente à Fig. 3. Além disso, Φ_n é a matriz PSD do ruído estacionário. Esta matriz PSD pode ser estimada por ex.: durante pausas de discurso. Se o ruído estacionário nos diferentes microfones for mutuamente independente, podemos simplesmente substituir Φ_n pela matriz de identidade da dimensão M pela M .

[072] Um esquema de blocos deste modelo é ilustrado na Fig. 4. Após transformação dos sinais de microfone com um banco de filtros (101), calculamos a matriz PSD Φ_n do ruído estacionário no bloco de estimativa de estatística de ruído (104). Além disso, calculamos a restrição linear de som difuso $\mathbf{b}_m$ no bloco (102) a partir do sinal ou utilizando à priori informação assumindo um campo sonoro difuso específico. No bloco (105), estimamos os DOAs das ondas planas L . A partir desta informação, calculamos as restrições de som direto $\mathbf{a}_l$ no bloco (106). A informação calculada é alimentada no bloco

de cálculo do filtro (103), que calcula as ponderações do filtro w_m com a solução em forma fechada apresentada em cima. Multiplicando as ponderações com os sinais de microfone rende a estimativa desejada do som difuso.

[073] Um exemplo de um padrão transmissor resultante para este filtro é indicado na Fig. 7. Aqui, os dois sons diretos chegam de uma direção azimutal de 51° e 97° . A figura ilustra o padrão transmissor resultante em uma frequência de 2,8 kHz quando utiliza uma rede linear uniforme com 16 microfones com espaçamento de microfones de 5 cm. O padrão transmissor possui nulos exatamente para 51° e 97° . Caso contrário, é quase omnidirecional. Isto é uma grande vantagem sobre o filtro espacial do estado da arte ilustrado na Fig. 2 que rende o filtro espacial na Fig. 6.

[074] O filtro calculado neste modelo possui as seguintes vantagens comparado com outro filtro espacial (por ex.: os filtros descritos na técnica anterior):

- Uma forte atenuação das ondas planas devido às restrições do som direto.
- Padrão transmissor quase omnidirecional desejado para captura de som difuso.

[075] Abordagem Combinada

[076] Os filtros espaciais ilustrados nas Figs. 3 e 4 fornecem geralmente um bom desempenho na prática. Contudo, também eles sofrem de inconvenientes específicos. Por exemplo, o filtro da Fig. 3 não suprime habitualmente na totalidade o som direto. A restante potência do som direto pode conduzir a efeitos indesejáveis durante reprodução de som espacial. Pelo contrário, o filtro espacial da Fig. 4 fornece uma robustez comparativamente fraca contra o ruído estacionário em baixas frequências. A seguir combinamos os filtros espaciais nas Figs. 3 e 4 de modo a explorar as vantagens de ambos filtros espaciais enquanto mitigam os inconvenientes. Em resultado disso, obtemos um filtro espacial caracterizado por uma boa supressão interferente

(supressão de som direto) e ruído estacionário de elevada robustez.

[077] A abordagem proposta a seguir representa um chamado filtro Wiener paramétrico multicanal (PMWF) que pode ser escalonado entre um chamado filtro espacial erro quadrático médio (MMSE) mínimo e o filtro espacial na Fig. 3. Adicionalmente, o filtro espacial na Fig. 4 é utilizado para estimar quantidades específicas exigidas.

[078] O vetor de ponderação do PMWF proposto é calculado como

$$\mathbf{w}_m(k, n) = \underset{\mathbf{w}}{\operatorname{argmin}} \mathbf{w}^H \Phi_x(k) \mathbf{w}$$

[079] sujeito a

$$E \left\{ \left| X_d(k, n, \mathbf{d}_m) - \hat{X}_d(k, n, \mathbf{d}_m) \right|^2 \right\} < \sigma^2$$

[080] em que σ^2 é o erro quadrático absoluto do som difuso estimado. A solução desta problema de otimização conduz a

$$\mathbf{w}_m(k, n) = \frac{\phi_d(k, n)}{\beta + 1} \Phi_x^{-1}(k, n) \mathbf{b}_m$$

[081] em que definimos

$$\beta = \alpha (\phi_d \mathbf{b}_m^H \Phi_x^{-1} \mathbf{b}_m)$$

[082] aqui, $\alpha \in [0, 1]$ é um parâmetro de controlo definido pelo utilizador. Para $\alpha = 0$, obtemos o filtro espacial MMSE que minimiza o erro quadrático médio do som difuso estimado. Para $\alpha = 1$, obtemos o filtro espacial proposto na Fig. 3. Para valores α entre eles, temos um escalonamento entre os dois filtros espaciais. Um α mais pequeno produz um melhor ruído e supressão interferente, enquanto um α mais elevado produz um som difuso estimado com menos distorção. De salientar que os elementos de $\mathbf{b}_m$ são calculados tal como explicado relativamente à Fig 3. A potência do som difuso ϕ_d pode ser calculada com o filtro espacial proposto na Fig. 4 que fornece uma supressão muito boa das ondas planas. Que $\mathbf{w}_1$ indique o filtro espacial na Fig. 4 para estimar o som difuso no primeiro microfone. Considerando $\mathbf{w}_1$, a potência do

som difuso no primeiro microfone pode por exemplo ser obtida tal como descrito em [Thiergart2013b], ou seja,

$$\hat{\phi}_d(k, n) = \frac{\mathbf{w}_1^H(k, n)[\Phi_x(k, n) - \Phi_n(k, n)]\mathbf{w}_1(k, n)}{\mathbf{w}_1^H(k, n)\Gamma_d\mathbf{w}_1(k, n)}$$

[083] ou com uma abordagem de decisão direta tal como explicado em [Kowalczyk2013]. Na formula anterior, Γ_d é a matriz de coerência espacial $M \times M$ para o som difuso. O (m, m') -th elemento de Γ_d é a coerência espacial $\gamma_{m, m'}$ entre microfone m e m' . Esta coerência espacial $\gamma_{m, m'}$ foi já definida em cima.

[084] Um esquema de blocos deste modelo é ilustrado na Fig. 5. Após transformação dos sinais de microfone com um banco de filtros (101), calculamos a matriz do sinal PSD Φ_x e a matriz de ruído PSD Φ_n no bloco (104). Além disso, calculamos a restrição linear do som difuso $\mathbf{b}_m$ no bloco (102) a partir do sinal ou utilizando à priori informação assumindo um campo sonoro difuso específico. No bloco (105), estimamos os DOAs das ondas planas L . A partir desta informação, calculamos as restrições de som direto $\mathbf{a}_l$ no bloco (106). Estas restrições são utilizadas em (107) juntamente com Φ_n para calcular as ponderações $\mathbf{w}_1$. A potência do som difuso ϕ_d é calculada em (108) a partir de $\mathbf{w}_1$ e Φ_n . As ponderações finais do filtro espacial $\mathbf{w}_m$ podem então ser calculadas em (103) utilizando ϕ_d , Φ_x , e $\mathbf{b}_m$. Com o parâmetro α pode-se escalonar o filtro espacial entre o filtro MMSE e PMWF. Multiplicando as ponderações $\mathbf{w}_m$ com os sinais de microfone rende a estimativa desejada do som difuso.

[085] Fig. 8 ilustra esquematicamente uma rede de microfones compreendendo M os microfones. A rede de microfones é exposta a um campo sonoro compreendendo partes de som direto e partes de som difuso. As partes de som direto são representadas na Fig. 8 por fontes sonoras L $X_{l=1}$ a $X_{l=L}$, cada uma assumindo como gerando uma onda significativamente plana na

localização da rede de microfones. A direção de propagação de cada onda plana (na localização da rede de microfones) é indicada na Fig. 8 pelos vetores $\mathbf{n}_1$ a $\mathbf{n}_L$. As partes do som direto são habitualmente uma função da localização $\mathbf{d}_m$. O som difuso $X_d(k, n, \mathbf{d}_m)$ pode modelar por exemplo a reverberação em um espaço. O som difuso é conhecido como sendo gerado por uma soma infinita de ondas planas com fases aleatórias, magnitude aleatória e DOAs aleatórias. Isto significa que o som difuso é gerado por um número infinito de fontes aleatoriamente distribuídas em redor da cena sonora. Estas cenas sonoras modelam o número infinito de reflexões do espaço que geram a última reverberação.

[086] Um sinal de individual de microfone, por exemplo para o m -th microfone, $X_m(k, n)$ é uma combinação das partes de som direto $\sum_{l=1}^L X_l(k, n, \mathbf{d}_m)$ para $X_{l=L}$, a parte de som difuso X_d , e ruído X_n , ou seja

$$X_m(k, n) = \sum_{l=1}^L X_l(k, n, \mathbf{d}_m) + X_d(k, n, \mathbf{d}_m) + X_n(k, n, \mathbf{d}_m)$$

[087] Funções de transferência relativa $B_{1,m}, B_{2,m}, \dots, B_{m',m}, \dots, B_{M,m}$ para o som difuso entre os outros microfones para o m -th microfone são esquematicamente ilustradas na Fig. 8. A função de transferência relativa $B_{m,m}$ do microfone m -th para si própria (não indicada na Fig. 8) é habitualmente é igual a 1.

[088] A lista seguinte fornece uma breve visão de alguns dos aspectos que foram agora descritos:

- Receber pelo menos dois sinais de microfone.
- Transformar os sinais de microfone no domínio de tempo-frequência ou outro domínio adequado.
- Calcular uma restrição linear de som difuso como uma função da correlação ou coerência do som difuso entre os microfones.
- Calcular estatísticas de sinal e/ou de ruído.
- Nalguns modelos: estimar os DOAs dos sons diretos e calcular restrições de som direto que representa as funções de transferência relativa

dos sons diretos entre os microfones.

- Nalguns modelos: calcular um filtro auxiliar e estimar a potência do som difuso.
- Calcular as ponderações de um filtro espacial para extrair o som difuso utilizando a estatística de sinal/ruído obtida e informação de potência de som difuso opcional considerando a restrição do som difuso.
- Executar uma combinação linear dos sinais de microfones utilizando as ponderações calculadas do filtro espacial.

[089] Apesar de alguns aspectos terem sido descritos no contexto de um aparelho, é claro que estes aspectos também representam uma descrição do correspondente método, em que um bloco ou dispositivo corresponde a um passo de método ou a uma característica de um passo de método. De modo análogo, os aspectos descritos no contexto de um passo de método também representam uma descrição de um correspondente bloco ou item ou característica de um correspondente aparelho.

[090] Dependendo de certos requisitos de implementação, os modelos da invenção podem ser implementados em hardware ou em software. A implementação pode ser executada utilizando um suporte de armazenamento digital, por exemplo um disquete, um DVD, um CD, uma ROM, uma PROM, uma EPROM, uma EEPROM ou uma memória FLASH, dotado de sinais de controlo lidos eletronicamente nela armazenados, que cooperam (ou são capazes de cooperarem) com um sistema de computador programável de modo que o respetivo método seja executado.

[091] Alguns modelos de acordo com a invenção compreendem um suporte de dados transitório dotado de sinais de controlo lidos eletronicamente, capazes de cooperarem com um sistema de computador programável, de modo que um dos métodos aqui descritos seja executado.

[092] Regra geral, modelos desta invenção podem ser implementados como um produto de programa de computador com um código de programa, sendo o código do programa operativo para executar um dos métodos quando

o produto do programa de computador é executado em um computador. O código de programa pode, por exemplo, ser guardado em um suporte de leitura em máquina.

[093] Outros modelos compreendem o programa de computador para executar um dos métodos aqui descritos, guardados em um suporte de leitura em máquina.

[094] Por outras palavras, um modelo do método da invenção é, por isso, um programa de computador com um código de programa para executar um dos métodos aqui descritos, quando o programa de computador corre em um computador.

[095] Outro modelo dos métodos da invenção é, por isso, um suporte de dados (ou um suporte de armazenamento digital ou um suporte de leitura em computador) compreendendo, aí gravados, o programa de computador para executar um dos métodos aqui descritos.

[096] Um modelo adicional do método inovador é, pois, um fluxo de dados ou uma sequência de sinais representando o programa de computador para executar um dos métodos aqui descritos. O fluxo de dados ou a sequência de sinais pode, por exemplo, ser configurado para ser transferido através de uma ligação de comunicação de dados, por exemplo via Internet.

[097] Um modelo adicional compreende um meio de processamento, por exemplo um computador, ou um dispositivo de lógica programável, configurado para ou adaptado para executar um dos métodos aqui descritos.

[098] Um modelo adicional compreende um computador tendo instalado nele o programa de computador para executar um dos métodos aqui descritos.

[099] Nalguns modelos, um dispositivo de lógica programável (por exemplo uma rede de portas lógicas programáveis) pode ser utilizado para executar algumas ou todas as funcionalidades dos métodos aqui descritos. Em alguns modelos, uma rede de portas lógicas programáveis pode cooperar com um microprocessador para executar um dos métodos aqui descritos. De um modo geral, os métodos são preferencialmente executados por qualquer

aparelho de hardware.

[100] Os modelos acima descritos são meramente ilustrativos para os princípios da presente invenção. Compreende-se que as modificações e variações das disposições e dos detalhes descritos serão evidentes aos profissionais da matéria. Pretende-se, por isso, que seja limitado apenas pelo âmbito das reivindicações impendentes da patente e não pelos detalhes específicos da descrição e explicação dos modelos aqui constantes.

[101] Referências

[Pulkki2007] V. Pulkki, "Spatial sound reproduction with directional audio coding," J. Audio Eng. Soc, Vol. 55, No. 6, pp. 503-516, June 2007.

[Thiergart2013] O. Thiergart, G. Del Galdo, M. Taseska, and E. Habets, "Geometry-based spatial sound acquisition using distributed microphone arrays," Audio, Speech, and Language Processing, IEEE Transactions on, Vol. 21, No. 12, pp. 2583-2594, December 2013.

[Kowalczyk2013] K. Kowalczyk, O. Thiergart, A. Craciun, and E. A. P. Habets, "Sound acquisition in noisy and reverberant environments using virtual", in Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013 IEEE Workshop on, October 2013.

[Thiergart2013b] O. Thiergart and E. A. P. Habets, "An informed LCMV filter based on multiple instantaneous direction-of-arrival estimates," in Acoustics Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013, pp. 659-663.

[Thiergart2012] O. Thiergart, G. D. Galdo, and E. A. P. Habets, "On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation," The Journal of the Acoustical Society of America, vol. 132, no. 4, pp. 2337– 2346, 2012.

[VanTrees2002] H. L. Van Trees, Detection, Estimation, and Modulation Theory: Part IV: Array Processing. John Wiley & Sons, April 2002, vol. 1.

[Elko2001] G. W. Elko, "Spatial coherence functions for differential

micro- phones in isotropic noise fields”, in Microphone Arrays: Signal Processing Techniques and Applications, edited by M. Brandstein and D. Ward, chapter 4, 61–85 (Springer, Berlin) (2001).

REIVINDICAÇÕES

1. Método para obtenção de coeficientes de filtro para um filtro de som difuso, o método **caracterizado por** compreender:

a definição de uma restrição linear para os coeficientes de filtro para o filtro de som difuso, sendo a restrição linear baseada em uma coerência espacial entre uma primeira parte de som difuso em um primeiro sinal de microfone e uma segunda parte de som difuso em um segundo sinal de microfone, sendo o primeiro sinal de microfone capturado por um primeiro microfone e o segundo sinal de microfone capturado por um segundo microfone afastado do primeiro microfone;

o cálculo de pelo menos uma de uma direção de chegada de pelo menos um som direto, estatística de sinal sobre o primeiro e o segundo sinal de microfone, e estatística de ruído sobre o primeiro e o segundo sinal de microfone;

a determinação de coeficientes de filtro do filtro de som difuso solucionando um problema de otimização relativo pelo menos a uma da direção de chegada do pelo menos um som direto, a estatística de sinal, e a estatística de ruído enquanto considera a restrição linear para os coeficientes de filtro; e

utilizando os coeficientes de filtro no filtro de som difuso para extrair um som difuso de ou para suprimir um som direto em uma gravação compreendendo o primeiro sinal de microfone e o segundo sinal de microfone, em que um ou mais entre a definição, o cálculo, a determinação, e a utilização é implementada, pelo menos em parte, por um ou mais elementos de hardware de um aparelho de processamento de sinal de áudio.

2. Método de acordo com a reivindicação 1, **caracterizado por** compreender

o fornecimento da coerência especial baseada em uma função de transferência relativa ou uma correlação do som difuso entre o primeiro microfone e o segundo microfone.

3. Método de acordo com a reivindicação 1, **caracterizado por** a coerência espacial ser baseada em uma medida anterior de uma função de transferência relativa ou uma correlação do som difuso para um dado ambiente durante períodos de tempo nos quais nenhum som direto está presente no ambiente.

4. Método de acordo com a reivindicação 1, **caracterizado por** a coerência espacial ser baseada em uma informação prévia assumindo um campo de som difuso específico tendo propriedades assumidas relativamente a uma correlação do som difuso entre o primeiro microfone eo segundo microfone.

5. Método de acordo com a reivindicação 1, **caracterizado por** o problema de otimização ser expresso por

$$\mathbf{w}_m(k, n) = \underset{\mathbf{w}}{\operatorname{argmin}} J(\mathbf{w})$$

estar sujeito à restrição linear

$$\mathbf{w}^H \mathbf{b}_m(k, n) = 1,$$

em que

$\mathbf{w}(k, n)$ é um vetor dos coeficientes de filtro do filtro de som difuso;

$\mathbf{w}_m(k, n)$ é a solução do problema de otimização baseado em uma avaliação de um sinal de microfone no microfone m-th;

$J(\mathbf{w})$ é uma função de custo;

$\mathbf{b}_m(k, n)$ é um vetor de coerências espaciais estimadas, em que o elemento m'-th do vetor é uma coerência espacial estimada do som difuso entre o microfone m-th e o microfone m'-th;

k é um índice do domínio de frequência; e

n é um índice do domínio de tempo.

6. Método de acordo com a reivindicação 5, **caracterizado por** a função de custo $J(\mathbf{w})$ ser baseada em uma das estatísticas de ruído, uma matriz de densidade espectral de potência, estatísticas de sinal, ou uma matriz de densidade espectral de potência do microfone.

7. Método de acordo com a reivindicação 1, **caracterizado por** compreender

a estimativa de pelo menos uma direção de chegada de pelo menos um som direto ou uma função de transferência relativa de pelo menos um som direto entre o primeiro microfone e o segundo microfone;

o cálculo de pelo menos uma restrição de som direto utilizando a direção de chegada da função de transferência relativa do pelo menos um som direto, por pelo menos uma restrição de som direto resultar em uma supressão do pelo menos um som direto.

8. Método de acordo com a reivindicação 1, **caracterizado por** uma solução ao problema de otimização ser

$$\mathbf{w}_m(k, n) = \frac{\phi_d(k, n)}{\beta + 1} \Phi_x^{-1}(k, n) \mathbf{b}_m$$

com

$$\beta = \alpha (\phi_d \mathbf{b}_m^H \Phi_x^{-1} \mathbf{b}_m)$$

em que

$\mathbf{w}_m(k, n)$ é uma solução do problema de otimização baseada em uma avaliação de um sinal de microfone no microfone m-th;

$\mathbf{b}_m(k, n)$ é um vetor de coerências espaciais estimadas, em que o elemento m'-th do vetor é uma coerência espacial estimada do som difuso entre o microfone m-th e o microfone m'-th;

$\alpha \in [0, 1]$ é um parâmetro de controlo definido pelo utilizador através do qual o filtro de som difuso pode ser escalonado entre um filtro espacial de erro quadrático médio mínimo e um filtro que minimize a potência de saída enquanto satisfaz a restrição de som difuso;

ϕ_d é uma potência de som difuso; e

Φ_x é uma matriz do espectro de potência dos sinais de microfone.

9. Método de acordo com a reivindicação 8, **caracterizado por**

compreender ainda

a estimativa da potência de som difuso ϕ_d com base em um filtro de som difuso auxiliar.

10. Método de acordo com a reivindicação 9, **caracterizado por** a estimativa da potência de som difuso ϕ_d ser executada com base em

$$\hat{\phi}_d(k, n) = \frac{\mathbf{w}_1^H(k, n) [\Phi_x(k, n) - \Phi_n(k, n)] \mathbf{w}_1(k, n)}{\mathbf{w}_1^H(k, n) \Gamma_d \mathbf{w}_1(k, n)}$$

em que

$\hat{\phi}_d$ é a potência de som difuso estimado que fornece uma estimativa para a potência de som difuso ϕ_d ;

$\mathbf{w}_1$ é um vetor de coeficientes de filtro do filtro de som difuso auxiliar;

Φ_x é uma matriz de densidade espectral de potência dos sinais de microfone;

Φ_n é uma matriz de densidade espectral de potência do ruído nos sinais de microfone; e

Γ_d é uma matriz de coerência espacial para o som difuso, o (m, m') -th elemento de Γ_d sendo uma coerência especial $\gamma_{m', m}$ entre o microfone m e m' .

11. Método de acordo com a reivindicação 1, **caracterizado por** compreender ainda:

em que a utilização dos coeficientes de filtro compreendem a execução de uma combinação linear do primeiro sinal de microfone e o segundo sinal de microfone utilizando os coeficientes de filtro do filtro de som difuso.

12. Método de acordo com a reivindicação 1, **caracterizado por** a utilização dos coeficientes de filtro no filtro de som difuso compreende multiplicar o sinal do primeiro microfone e o sinal do segundo microfone pelos coeficientes de filtro para obter uma estimativa de som difuso na gravação.

13. Aparelho para obtenção de coeficientes de filtro para um filtro de som difuso, o aparelho **caracterizado por** compreender:

um calculador de restrição linear configurado para definir uma restrição linear para os coeficientes de filtro do filtro de som difuso, sendo a restrição linear baseada em uma coerência espacial entre uma primeira parte de som difuso em um primeiro microfone e uma segunda parte de som difuso em um segundo microfone, sendo o primeiro sinal de microfone capturado por um primeiro microfone e o segundo sinal de microfone capturado por um segundo microfone afastado do primeiro microfone;

um calculador configurado para calcular pelo menos um de uma direção de chegada de pelo menos um som direto, estatística de sinal sobre o primeiro e o segundo sinal de microfone, e estatística de ruído sobre o primeiro e segundo sinal de microfone; e

um calculador de coeficientes de filtro configurado para determinar os coeficientes de filtro do filtro de som difuso solucionando um problema de otimização relativa a pelo menos um da direção de chegada de pelo menos um som direto, a estatística de sinal e a estatística de ruído enquanto considera a restrição linear para os coeficientes de filtro,

em que o aparelho está configurado para utilizar os coeficientes de filtro no filtro de som difuso para extrair um som difuso de ou para suprimir um som direto em uma gravação compreendendo o primeiro sinal de microfone e o segundo sinal de microfone, e

em que um ou mais entre a calculadora de restrições lineares, a calculadora, e a calculadora de coeficientes de filtro é implementada, pelo menos em parte, por um ou mais elementos de hardware do aparelho.

14. Aparelho de acordo com a reivindicação 13, **caracterizado por** a coerência espacial ser baseada em uma função de transferência relativa do som difuso entre o primeiro microfone e o segundo microfone.

15. Aparelho de acordo com a reivindicação 13, **caracterizado por** compreender ainda

um calculador de coeficientes de filtro auxiliar configurado para determinar coeficientes de filtro auxiliar de um filtro de som difuso auxiliar solucionando um diferente problema de otimização enquanto considera uma diferente restrição linear para os coeficientes de filtro auxiliar do que o calculador de coeficientes de filtro;

o filtro de som difuso auxiliar configurado para estimar uma potência de som difuso estimado;

em que o calculador dos coeficientes de filtro é configurado para ter em consideração a potência de som difuso estimado quando determina os coeficientes de filtro do filtro de som difuso.

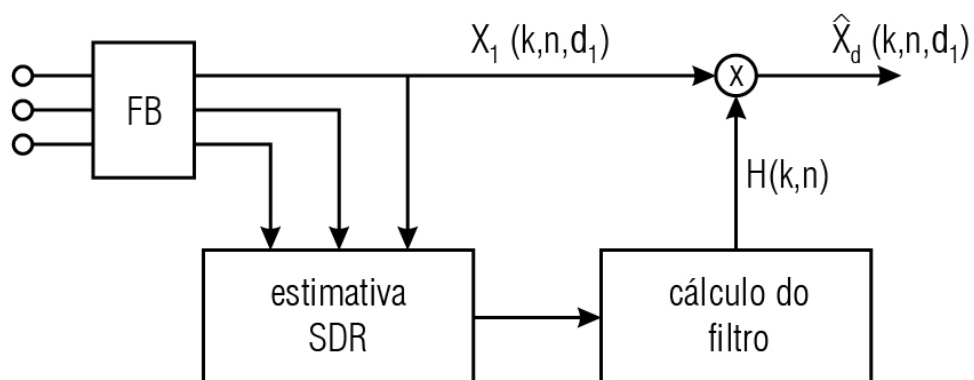


FIGURA 1

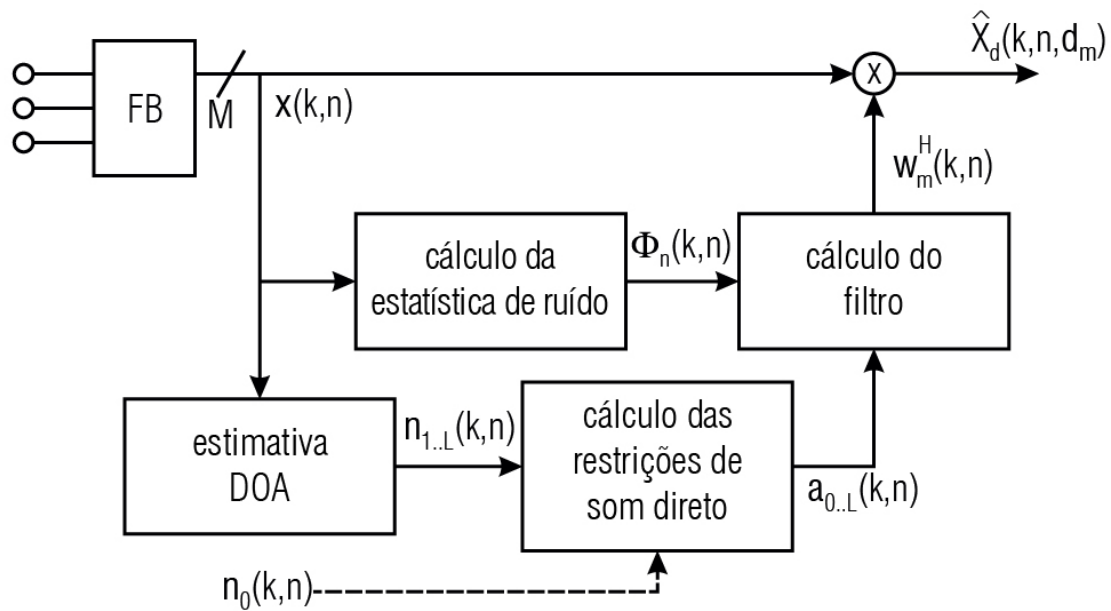


FIGURA 2

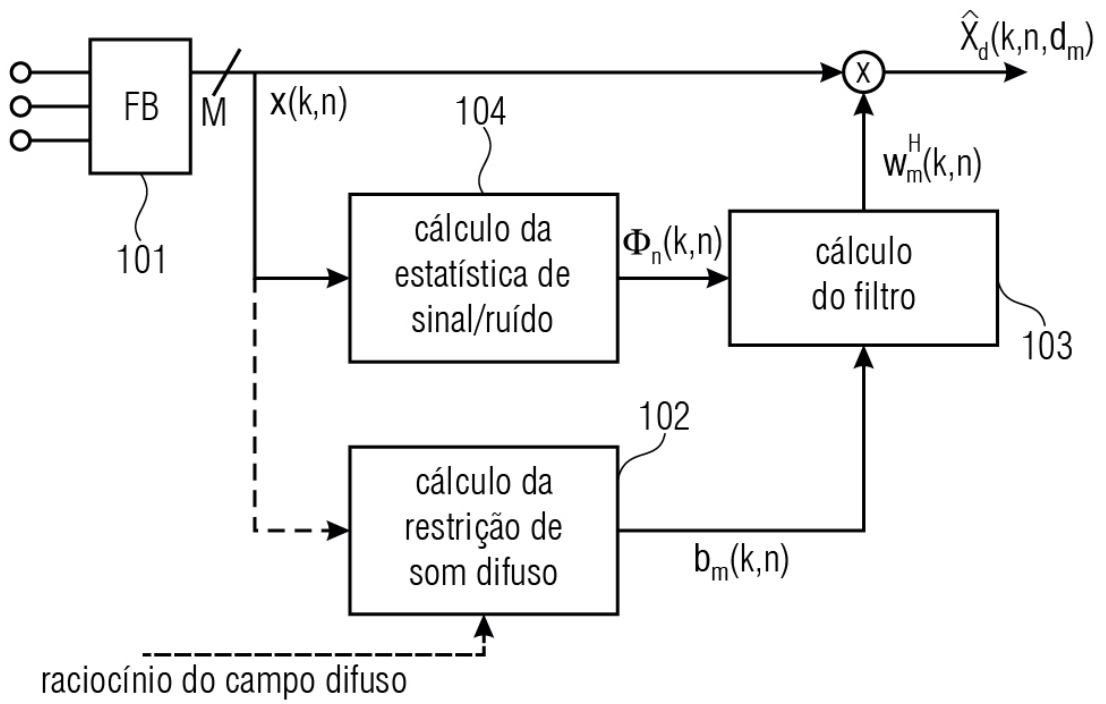


FIGURA 3

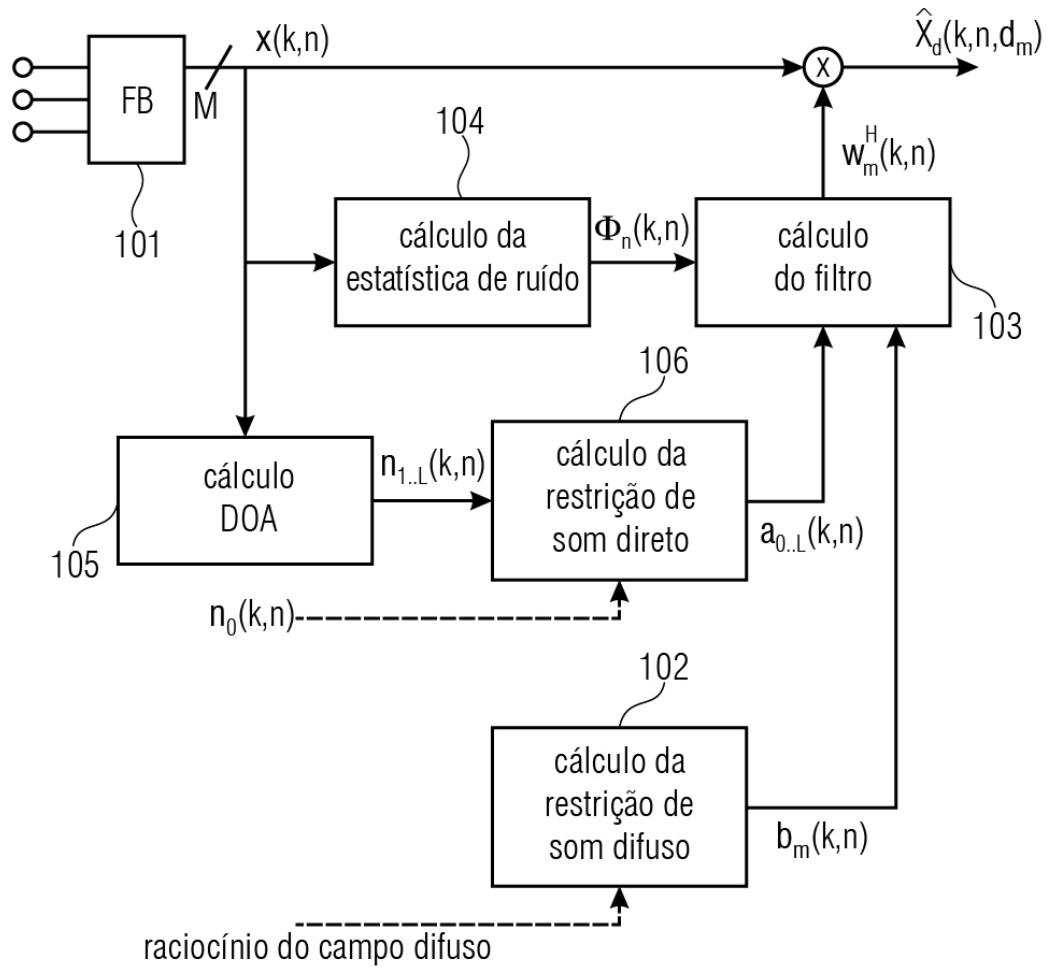


FIGURA 4

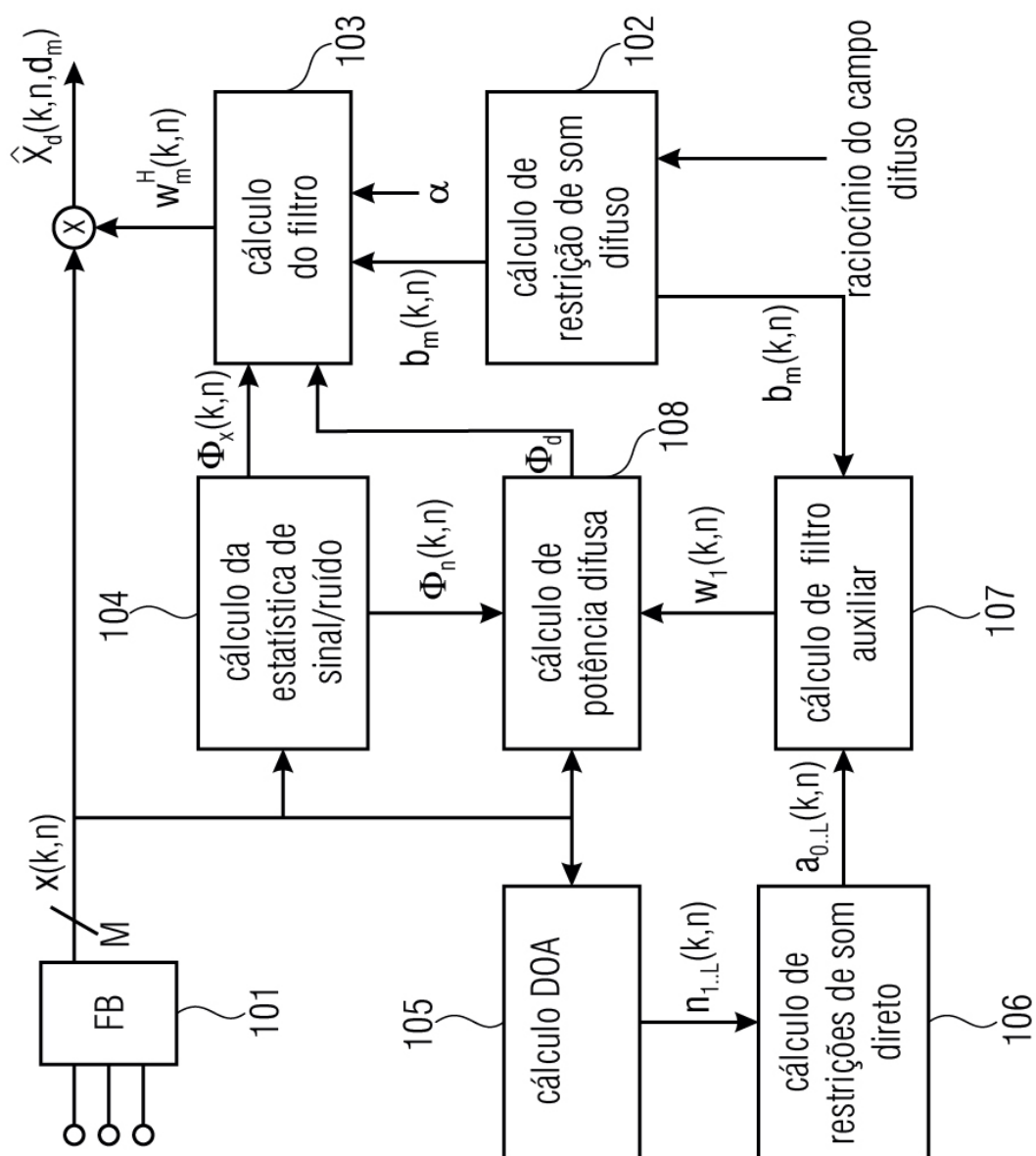


FIGURA 5

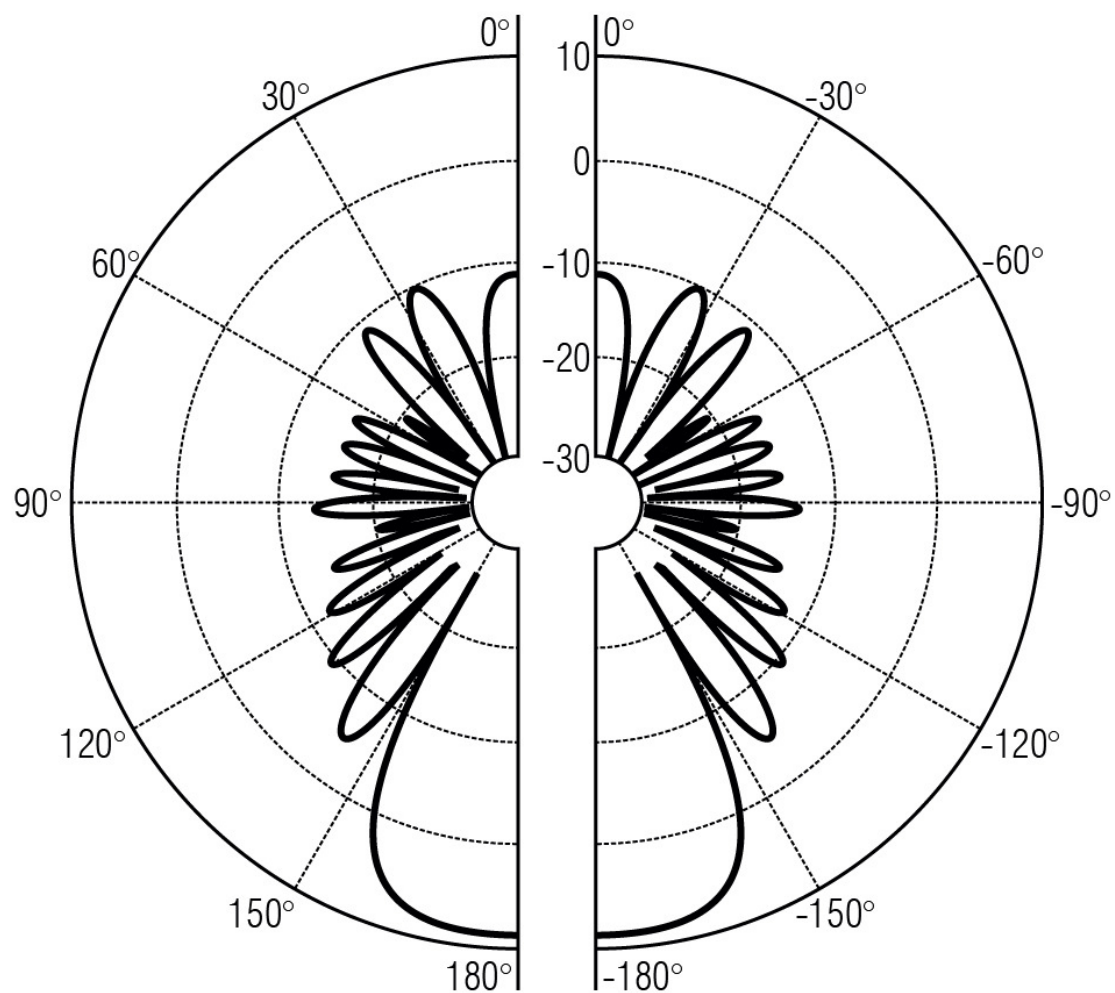


FIGURA 6

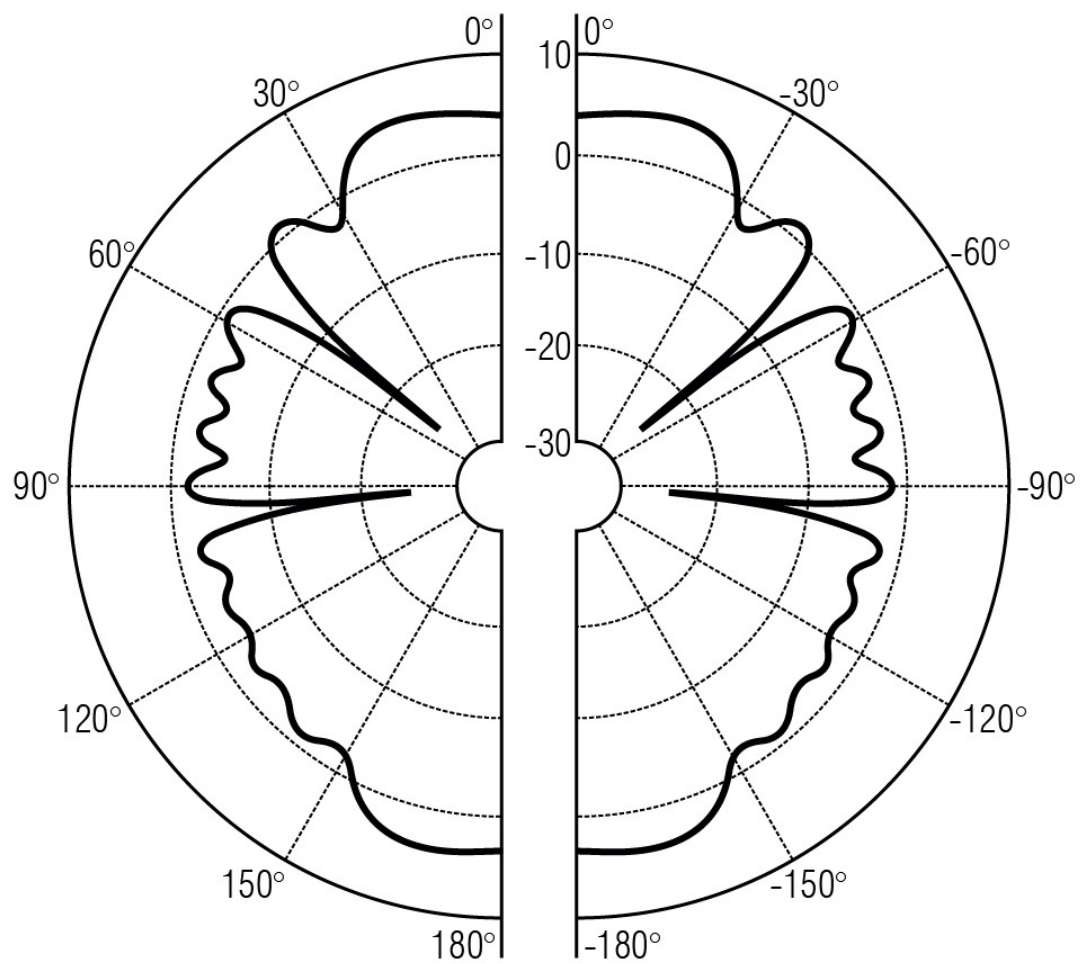


FIGURA 7

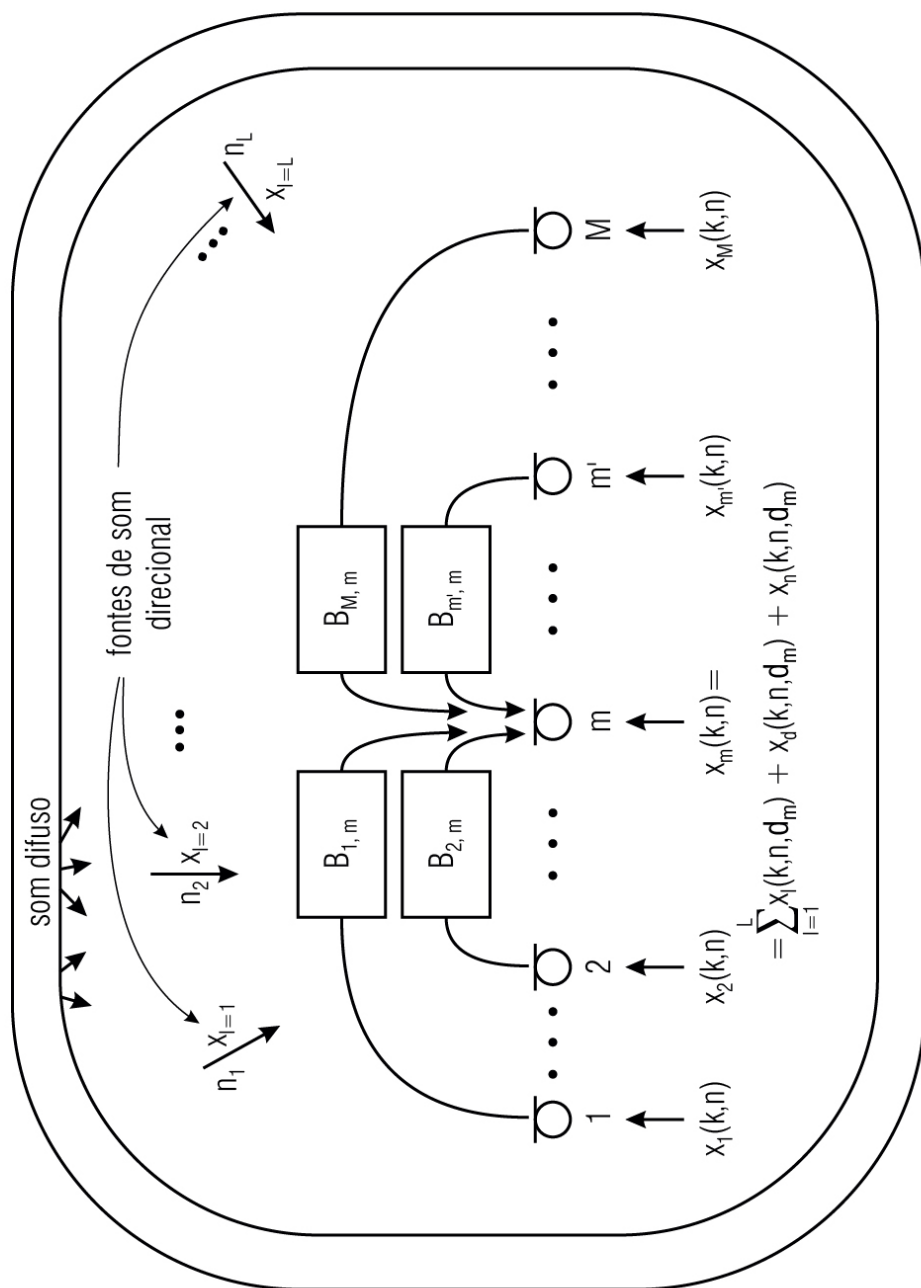


FIGURA 8