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

(11) BR 112016025771-5 B1

(22) Data do Depósito: 23/04/2015

(45) Data de Concessão: 23/08/2022

(54) Título: SISTEMA, APARELHO E MÉTODO PARA REPRODUÇÃO DE CENA ACÚSTICA CONSISTENTE COM BASE NA FILTRAGEM ESPACIAL INFORMADA

(51) Int.Cl.: H04S 7/00; H04R 25/00.

(30) Prioridade Unionista: 05/05/2014 EP 14167053.9; 05/09/2014 EP 14183855.7.

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

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

(86) Pedido PCT: PCT EP2015058859 de 23/04/2015

(87) Publicação PCT: WO 2015/169618 de 12/11/2015

(85) Data do Início da Fase Nacional: 03/11/2016

(57) Resumo: SISTEMA, APARELHO E MÉTODO PARA REPRODUÇÃO DE CENA ACÚSTICA CONSISTENTE COM BASE NA FILTRAGEM ESPACIAL INFORMADA. Sistema para gerar um ou mais sinais de saída de áudio é fornecido. O sistema compreende um módulo de decomposição (101), um processador de sinais (105), e uma interface de saída (106). O módulo de decomposição (101) é configurado para receber dois ou mais sinais de entrada de áudio, em que o módulo de decomposição (101) é configurado para gerar um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de entrada de áudio, e em que o módulo de decomposição (101) é configurado para gerar um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de entrada de áudio. O processador de sinais (105) é configurado para receber o sinal de componente direto, o sinal de componente difuso e informações de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio. Além disso, o processador de sinais (105) é configurado para gerar um ou mais sinais difusos processados dependendo do sinal de componente difuso. Para cada sinal de saída de áudio de um ou (...).

Relatório Descritivo da Patente de Invenção: **"SISTEMA, APARELHO E MÉTODO PARA REPRODUÇÃO DE CENA ACÚSTICA CONSISTENTE COM BASE NA FILTRAGEM ESPACIAL INFORMADA".**

[001] Descrição

[002] A presente invenção se refere ao processamento de sinais de áudio, e, em particular, a um sistema, um aparelho e um método para a reprodução de cena acústica consistente com base na filtragem espacial informada.

[003] Na reprodução de som espacial o som no local de gravação (lado de extremidade próxima) é capturado com múltiplos microfones e, em seguida, reproduzido no lado da reprodução (lado de extremidade distante) usando múltiplos alto-falantes ou fones de ouvido. Em muitas aplicações, é desejável reproduzir o som gravado de modo que a imagem espacial recriada no lado de extremidade distante seja consistente com a imagem espacial original no lado de extremidade próxima. Isto significa, por exemplo, que o som das fontes de som é reproduzido a partir das direções onde as fontes estavam presentes no cenário de gravação original. Em alternativa, quando, por exemplo, um vídeo está complementando o áudio gravado, é desejável que o som seja reproduzido de modo que a imagem acústica recriada seja consistente com a imagem de vídeo. Isto significa, por exemplo, que o som de uma fonte de som é reproduzido a partir da direção em que a fonte é visível no vídeo. Além disso, a câmera de vídeo pode ser equipada com uma função de zoom visual ou o usuário no lado de extremidade distante pode aplicar um zoom digital para o vídeo, o que pode mudar a imagem visual. Neste caso, a imagem acústica do som espacial reproduzido deve mudar em conformidade. Em muitos casos, o lado de extremidade distante determina a imagem espacial com a qual o som reproduzido deve ser consistente e é determinada quer no lado de extremidade distante ou durante a reprodução, por exemplo, quando uma imagem de vídeo está envolvida. Consequentemente, o som espacial no lado de extremidade próxima deve ser registrado, processado e transmitido de modo que no lado de extremidade distante ainda podemos controlar a imagem acústica recriada.

[004] A possibilidade de reproduzir uma cena acústica gravada de forma consistente com uma imagem espacial desejada é necessária em muitas aplicações modernas. Por exemplo, dispositivos de consumo modernos, tais

como câmeras digitais ou telefones celulares são muitas vezes equipados com uma câmera de vídeo e vários microfones. Isso permite gravar vídeos juntamente com o som espacial, por exemplo, o som estéreo. Quando se reproduz o áudio gravado em conjunto com o vídeo, é desejável que a imagem acústica e visual sejam consistentes. Quando o usuário aumenta o zoom com a câmera, é desejável para recriar o efeito de zoom visual acusticamente de modo que as imagens acústicos e visuais estejam alinhadas ao assistir o vídeo. Por exemplo, quando o usuário aumenta o zoom em uma pessoa, a voz desta pessoa deve se tornar menos reverberante à medida que a pessoa parece estar mais perto da câmera. Além disso, a voz da pessoa deve ser reproduzida a partir da mesma direção em que a pessoa aparece na imagem visual. A imitação do zoom visual de uma câmera acusticamente é chamada de zoom acústico a seguir e representa um exemplo de uma reprodução de áudio e vídeo consistente. A reprodução de áudio e vídeo consistente que pode envolver um zoom acústico também é útil em teleconferências, onde o som espacial no lado de extremidade próxima é reproduzido no lado de extremidade distante em conjunto com uma imagem visual. Além disso, é desejável recriar o efeito de zoom visual acusticamente de modo que as imagens visuais e acústicas estejam alinhadas.

[005] A primeira implementação de um zoom acústico foi apresentada em [1], em que o efeito de zoom foi obtido aumentando a diretividade de um microfone direcional de segunda ordem, cujo sinal foi gerado com base nos sinais de uma disposição linear de microfones. Esta abordagem foi estendida em [2] para um zoom estéreo. Uma abordagem mais recente para um zoom mono ou estéreo foi apresentada em [3], que consiste em alterar os níveis de fonte de som de modo que a fonte a partir da direção frontal é conservada, enquanto que as fontes provenientes de outras direções e o som difuso foram atenuadas. As abordagens propostas em [1,2] resultam em um aumento da razão de direto-para-reverberação (direct-to-reverberation ratio, DRR) e a abordagem em [3] permite, adicionalmente, a supressão de fontes indesejáveis. As abordagens acima mencionadas assumem que a fonte de som está localizada na frente de uma câmera, e não têm o objetivo de capturar a imagem acústica que é consistente com a imagem de vídeo.

[006] Uma abordagem bem conhecida para uma gravação e reprodução de

som espacial flexível é representada pela codificação de áudio direcional (directional audio coding, DirAC) [4]. Em DirAC, o som espacial no lado de extremidade próxima é descrito em termos de um sinal de áudio e informações laterais paramétricas, a saber, a direção de chegada (direction-of-arrival, DOA) e dispersividade do som. A descrição paramétrica permite a reprodução da imagem espacial original com configurações de alto-falante arbitrárias. Isto significa que a imagem espacial recriada no lado de extremidade distante é consistente com a imagem espacial durante a gravação no lado de extremidade próxima. No entanto, se, por exemplo, um vídeo está complementando o áudio gravado, então, o som espacial reproduzido não está necessariamente alinhado com a imagem de vídeo. Além disso, a imagem acústica recriada não pode ser ajustada quando as imagens visuais se alteram, por exemplo, quando a direção do olhar e o zoom da câmera são alterados. Isto significa que DirAC não fornece qualquer possibilidade para ajustar a imagem acústica recriada para uma imagem espacial desejada arbitrária.

[007] Em [5], um zoom acústico foi realizado com base em DirAC. DirAC representa uma base razoável para realizar um zoom acústico à medida que se baseia em um modelo de sinal simples, mas poderoso assumindo que o campo sonoro no domínio do tempo-frequência é composto de uma única onda plana e de som difuso. Os parâmetros do modelo subjacentes, por exemplo, o DOA e a dispersividade, são explorados para separar o som direto e o som difuso e para criar o efeito de zoom acústico. A descrição paramétrica do som espacial permite uma transmissão eficiente da cena sonora para o lado de extremidade distante enquanto continua a fornecer ao usuário o controle total sobre o efeito de zoom e reprodução de som espacial. Apesar de DirAC empregar vários microfones para estimar os parâmetros de modelo, apenas filtros de canal único são aplicados para extrair o som direto e o som difuso, o que limita a qualidade do som reproduzido. Além disso, todas as fontes na cena de som são assumidas como estando posicionadas sobre um círculo e a reprodução de som espacial é realizada com referência a uma alteração da posição de uma câmara áudio-visual, o que é inconsistente com o zoom visual. Na verdade, o zoom altera o ângulo de visão da câmara, enquanto a distância para os objetos visuais e as suas posições relativas à imagem permanecem inalteradas, o que está em

contraste com a movimentação da câmera.

[008] Uma abordagem relacionada é a chamada técnica de microfone virtual (virtual microphone, VM) [6,7], que considera o mesmo modelo de sinal que DirAC mas permite sintetizar o sinal de um microfone não existente (virtual) em uma posição arbitrária na cena sonora. O movimento de VM em direção a uma fonte de som é análogo ao movimento da câmera para uma nova posição. O VM foi realizado usando filtros de multicanais para melhorar a qualidade do som, mas requer vários conjuntos de microfones distribuídos para estimar os parâmetros do modelo.

[009] No entanto, seria altamente apreciada, se conceitos adicionais melhorados para processamento de sinais de áudio fossem fornecidos.

[010] Desse modo, o objetivo da presente invenção consiste em fornecer conceitos melhorados para processamento de sinais de áudio. O objetivo da presente invenção é resolvido por um sistema de acordo com a reivindicação 1, por um aparelho de acordo com a reivindicação 13, por um método de acordo com a reivindicação 14, por um método de acordo com a reivindicação 15 e por um programa de computador de acordo com a reivindicação 16.

[011] Um sistema para gerar um ou mais sinais de saída de áudio é fornecido. O sistema compreende um módulo de decomposição, um processador de sinais, e uma interface de saída. O módulo de decomposição é configurado para receber dois ou mais sinais de entrada de áudio, em que o módulo de decomposição é configurado para gerar um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de entrada de áudio, e em que o módulo de decomposição é configurado para gerar um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de entrada de áudio. O processador de sinais é configurado para receber o sinal de componente direto, o sinal de componente difuso e informações de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio. Além disso, o processador de sinais é configurado para gerar um ou mais sinais difusos processados dependendo do sinal de componente de desarme. Para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, o processador de sinais é configurado para determinar,

dependendo da direção de chegada, um ganho direto, o processador de sinais é configurado para aplicar dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e o processador de sinais é configurado para combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio. A interface de saída é configurada para produzir o um ou mais sinais de saída de áudio.

[012] De acordo com modalidades, os conceitos são fornecidos para obtenção de gravação e reprodução de som espacial de tal modo que a imagem acústica recriada pode, por exemplo, ser consistente com uma imagem espacial desejada, que é, por exemplo, determinada pelo usuário, no lado de extremidade distante ou por uma imagem de vídeo. A abordagem proposta usa uma matriz de microfone no lado de extremidade próxima que nos permite decompor o som captado em componentes som direto e um componente de som difuso. Os componentes de som extraídos são então transmitidos para o lado de extremidade distante. A reprodução de som espacial consistente pode, por exemplo, ser realizada por uma soma ponderada do som direto extraído e do som difuso, em que as ponderações dependem da imagem espacial desejada para a qual o som reproduzido deve ser consistente, por exemplo, as ponderações dependem da direção do olhar e do fator de zoom da câmera de vídeo, que pode, por exemplo, estar complementando a gravação de áudio. Os conceitos são fornecidos os quais empregam filtros de multicanais informados para a extração de som direto e de som difuso.

[013] De acordo com uma modalidade, o processador de sinais pode, por exemplo, ser configurado para determinar dois ou mais sinais de saída de áudio, em que para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio, uma função de ganho de panning de imagem pode, por exemplo, ser atribuída ao dito sinal de saída de áudio, em que a função de ganho de panning de cada um de dois ou mais sinais de saída de áudio compreende uma pluralidade de valores de argumento de função de panning, em que um valor de retorno de função de panning pode, por exemplo, ser atribuído a cada um dos ditos valores de argumento de função de panning, em que, quando dita função de ganho de panning recebe um dos ditos valores de argumento de função de panning, dita função de ganho de panning pode, por exemplo, ser configurada

para retornar o valor de retorno de função de panning sendo atribuído ao dito um dos ditos valores de argumento de função de panning, e em que o processador de sinais pode, por exemplo, ser configurado para determinar cada um de dois ou mais sinais de saída de áudio dependendo de um valor de argumento dependente de direção dos valores de argumento de função de panning da função de ganho de panning sendo atribuída ao dito sinal de saída de áudio, em que o dito valor de argumento dependente de direção depende da direção de chegada.

[014] Em uma modalidade, a função de ganho de panning de cada um de dois ou mais sinais de saída de áudio tem um ou mais máximos globais, sendo um dos valores de argumento de função de panning, em que para cada um de um máximo global de cada função de ganho de panning, nenhum outro valor de argumento de função de panning existe para a qual a dita função de ganho de panning retorna um valor de retorno de função de panning maior do que para o dito máximo global, e em que, para cada par de um primeiro sinal de saída de áudio e um segundo sinal de saída de áudio de dois ou mais de sinais de saída de áudio, pelo menos um de um ou mais máximos globais da função de ganho de panning do primeiro sinal de saída de áudio pode, por exemplo, ser diferente de qualquer um de um ou mais máximos globais da função de ganho de panning do segundo sinal de saída de áudio.

[015] De acordo com uma modalidade, o processador de sinais pode, por exemplo, ser configurado para gerar cada sinal de saída de áudio de um ou mais sinais de saída de áudio dependendo de uma função de janela de ganho, em que a função de janela de ganho pode, por exemplo, ser configurada para retornar um valor de retorno de função de janela quando receber um valor de argumento de função de janela, em que, se o valor de argumento de função de janela pode, por exemplo, ser maior do que um limite de janela inferior e menor do que um limite de janela superior, a função de janela de ganho pode, por exemplo, ser configurada para retornar um valor de retorno de função de janela sendo maior do que qualquer valor de retorno de função de janela retornado pela função de janela de ganho, se o valor de argumento de função de janela puder, por exemplo, ser menor do que o limite inferior, ou maior do que o limite superior.

[016] Em uma modalidade, o processador de sinais pode, por exemplo, ser

configurado para receber, adicionalmente, informações de orientação, indicando um desvio angular de uma direção de busca em relação à direção de chegada, e em que pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende das informações de orientação; ou em que o módulo de computação de função de ganho pode, por exemplo, ser configurado para receber informações de zoom, em que as informações de zoom indicam um ângulo de abertura de uma câmara, e em que pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende das informações de zoom; ou em que o módulo de computação de função de ganho pode, por exemplo, ser configurado para receber, adicionalmente, um parâmetro de calibração, e em que pelo menos uma dentre a função do ganho de panning e a função de janela de ganho depende do parâmetro de calibração.

[017] De acordo com uma modalidade, o processador de sinais pode, por exemplo, ser configurado para receber informações de distância, em que o processador de sinais pode, por exemplo, ser configurado para gerar cada sinal de saída de áudio de um ou mais sinais de saída de áudio, dependendo das informações de distância.

[018] De acordo com uma modalidade, o processador de sinais pode, por exemplo, ser configurado para receber um valor de ângulo original, dependendo de uma direção de chegada original, sendo a direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio, e pode, por exemplo, ser configurado para receber as informações de distância, em que o processador de sinais pode, por exemplo, ser configurado para calcular um valor de ângulo modificado dependendo do valor de ângulo original e dependendo das informações de distância, e em que o processador de sinais pode, por exemplo, ser configurado para gerar cada sinal de saída de áudio de um ou mais sinais de saída de áudio dependendo do valor de ângulo modificado.

[019] De acordo com uma modalidade, o processador de sinais pode, por exemplo, ser configurado para gerar um ou mais sinais de saída de áudio através da condução de filtragem passa-baixo, ou através da adição de som direto retardado, ou através da condução de atenuação do som direto, ou através da condução de suavização temporal, ou através da condução de direção de dispersão de chegada, ou através da condução de descorrelação.

[020] Em uma modalidade, o processador de sinais pode, por exemplo, ser configurado para gerar dois ou mais canais de saída de áudio, em que o processador de sinais pode, por exemplo, ser configurado para aplicar o ganho difuso sobre o sinal de componente difuso para obter um sinal difuso intermediário, e em que o processador de sinais pode, por exemplo, ser configurado para gerar um ou mais sinais descorrelacionados a partir do sinal difuso intermediário através da condução de descorrelação, em que o um ou mais sinais descorrelacionados formam um ou mais sinais de difusos processados, ou em que o sinal difuso intermediário e o um ou mais sinais descorrelacionados formam o um ou mais sinais difusos processados.

[021] De acordo com uma modalidade, o sinal de componente direto e um ou mais sinais de componente diretos formam um grupo de dois ou mais sinais de componente diretos, em que o módulo de decomposição pode, por exemplo, ser configurado para gerar um ou mais sinais de componente diretos adicionais compreendendo, adicionalmente, componentes de sinal direto de dois ou mais sinais de entrada de áudio, em que a direção de chegada e uma ou mais direções de chegadas adicionais forma um grupo de duas ou mais direções de chegadas, em que cada direção de chegada do grupo de duas ou mais direções de chegadas pode, por exemplo, ser atribuída a exatamente um sinal de componente direto do grupo de dois ou mais sinais de componente diretos, em que o número de sinais de componente diretos de dois ou mais sinais de componente diretos e o número de direção de chegadas das duas direções de chegada pode, por exemplo, ser igual, em que o processador de sinais pode, por exemplo, ser configurado para receber o grupo de dois ou mais sinais de componente diretos, e o grupo de duas ou mais direções de chegadas, e em que, para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, o processador de sinais pode, por exemplo, ser configurado para determinar, para cada sinal de componente direto do grupo de dois ou mais sinais de componente diretos, um ganho direto, dependendo da direção de chegada do dito sinal de componente direto, o processador de sinais pode, por exemplo, ser configurado para gerar um grupo de dois ou mais sinais diretos processados através da aplicação, para cada sinal de componente direto do grupo de dois ou mais sinais de componente diretos, o ganho direto do dito sinal de componente

direto sobre o dito sinal de componente direto, e o processador de sinais pode, por exemplo, ser configurado para combinar um de um ou mais sinais difusos processados e cada sinal processado do grupo de dois ou mais sinais processados para gerar o dito sinal de saída de áudio.

[022] Em uma modalidade, o número dos sinais de componente diretos do grupo de dois ou mais sinais de componente diretos e 1 pode, por exemplo, ser menor do que o número de sinais de entrada de áudio que é recebido pela interface de recepção.

[023] Além disso, um aparelho auditivo ou um dispositivo de audição auxiliar compreendendo um sistema tal como acima descrito pode, por exemplo, ser fornecido.

[024] Além disso, um aparelho para gerar um ou mais sinais de saída de áudio é fornecido. O aparelho compreende um processador de sinais e uma interface de saída. O processador de sinais é configurado para receber um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de áudio originais, em que o processador de sinais é configurado para receber um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de áudio originais, e em que o processador de sinais é configurado para receber informação de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio. Além disso, o processador de sinais é configurado para gerar um ou mais sinais difusos processados dependendo do sinal de componente de desarme. Para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, o processador de sinais é configurado para determinar, dependendo da direção de chegada, um ganho direto, o processador de sinais é configurado para aplicar dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e o processador de sinais é configurado para combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio. A interface de saída é configurada para produzir o um ou mais sinais de saída de áudio.

[025] Além disso, um método para gerar um ou mais sinais de saída de áudio é fornecido. O método compreende:

[026] - Receber dois ou mais sinais de entrada de áudio.

[027] - Gerar um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de entrada de áudio.

[028] - Gerar um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de entrada de áudio.

[029] - Receber informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio.

[030] - Gerar um ou mais sinais difusos processados dependendo do sinal de componente de desarme.

[031] - Para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, determinar, dependendo da direção de chegada, um ganho direto, aplicar o dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio. E:

[032] - Produzir o um ou mais sinais de saída de áudio.

[033] Além disso, um método para gerar um ou mais sinais de saída de áudio é fornecido. O método compreende:

[034] - Receber um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de áudio originais.

[035] - Receber um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de áudio originais.

[036] - Receber informações de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio.

[037] - Gerar um ou mais sinais difusos processados, dependendo do sinal de componente de desarme.

[038] - Para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, determinar, dependendo da direção de chegada, um ganho direto, aplicar o dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio. E:

[039] - Produzir o um ou mais sinais de saída de áudio.

[040] Além disso, os programas de computador são fornecidos, em que cada um dos programas de computador é configurado para implementar um dos métodos acima descritos quando estiver sendo executado em um computador ou processador de sinais, de modo que cada um dos métodos acima descrito é implementado por um dos programas de computador.

[041] Além disso, um sistema para gerar um ou mais sinais de saída de áudio é fornecido. O sistema compreende um módulo de decomposição, um processador de sinais, e uma interface de saída. O módulo de decomposição é configurado para receber dois ou mais sinais de entrada de áudio, em que o módulo de decomposição é configurado para gerar um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de entrada de áudio, e em que o módulo de decomposição é configurado para gerar um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de entrada de áudio. O processador de sinais é configurado para receber o sinal de componente direto, o sinal de componente difuso e informações de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio. Além disso, o processador de sinais é configurado para gerar um ou mais sinais difusos processados dependendo do sinal de componente de desarme. Para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, o processador de sinais é configurado para determinar, dependendo da direção de chegada, um ganho direto, o processador de sinais é configurado para aplicar dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e o processador de sinais é configurado para combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio. A interface de saída é configurada para produzir o um ou mais sinais de saída de áudio. O processador de sinais compreende um módulo de computação de função de ganho para calcular uma ou mais funções de ganho, em que cada função de ganho de uma ou mais funções de ganho, compreende uma pluralidade de valores de argumento de função de ganho, em que um valor de retorno de função de ganho é atribuído a cada um dos ditos valores de argumento de função de ganho, em que, quando a dita função de ganho recebe um dos ditos valores de argumento

de função de ganho, em que a dita função de ganho é configurada para retornar o valor de retorno de função de ganho que está sendo atribuído a um dos ditos valores de argumento de função de ganho. Além disso, o processador de sinais compreende, adicionalmente, um modificador de sinais para selecionar, dependendo da direção de chegada, um valor de argumento dependente de direção a partir dos valores de argumento de função de ganho de uma função de ganho de uma ou mais funções de ganho, para obter o valor de retorno de função de ganho que é atribuído ao dito valor de argumento dependente de direção a partir da dita função de ganho, e para determinar o valor de ganho de pelo menos um de um ou mais sinais de saída de áudio, dependendo do dito valor de retorno de função de ganho obtido a partir da dita função de ganho.

[042] De acordo com uma modalidade, o módulo de computação de função de ganho pode, por exemplo, ser configurado para gerar uma tabela de pesquisa para cada função de ganho de uma ou mais funções de ganho, em que a tabela de pesquisa compreende uma pluralidade de entradas, em que cada uma das entradas da tabela de pesquisa compreende um dentre os valores de argumento de função de ganho e o valor de retorno de função de ganho que está sendo atribuído ao dito valor de argumento de função de ganho, em que o módulo de computação de função ganho pode, por exemplo, ser configurado para armazenar a tabela de pesquisa de cada função ganho na memória persistente ou não persistente, e em que o modificador de sinais pode, por exemplo, ser configurado para obter o valor da função de ganho de retorno sendo atribuído ao dito valor de argumento dependente de direção através da leitura do dito valor de retorno de ganho de função a partir de uma ou mais tabelas de pesquisa que são armazenadas na memória.

[043] Em uma modalidade, o processador de sinais pode, por exemplo, ser configurado para determinar dois ou mais sinais de saída de áudio, em que o módulo de computação de função de ganho pode, por exemplo, ser configurado para calcular duas ou mais funções de ganho, em que, para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio, o módulo de computação de função de ganho pode, por exemplo, ser configurado para calcular uma função de ganho de panning sendo atribuída ao dito sinal de saída de áudio de uma das duas ou mais funções de ganho, em que o modificador de sinais pode, por

exemplo, ser configurado para gerar o dito sinal de saída de áudio dependendo da dita função de ganho de panning.

[044] De acordo com uma modalidade, a função de ganho de panning de cada um dos dois ou mais sinais de saída de áudio pode, por exemplo, ter um ou mais máximos globais, sendo um dos valores de argumento de função de ganho da dita função de ganho de panning, em que para cada um de um ou mais máximos globais da dita função de ganho de panning, não existe nenhum outro valor de argumento de função de ganho para o qual a dita função de ganho de panning retorna um valor de retorno de função de ganho maior do que para o dito máximo global, e em que, para cada par de um primeiro sinal de saída de áudio e um segundo sinal de saída de áudio de dois ou mais sinais de saída de áudio, pelo menos, um de um ou mais máximos globais da função de ganho de panning do primeiro sinal de saída de áudio poderá, por exemplo, ser diferente de qualquer um de um ou mais máximos globais da função de ganho de panning do segundo sinal de saída de áudio.

[045] De acordo com uma modalidade, para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio, o módulo de computação de função de ganho pode, por exemplo, ser configurado para calcular uma função de janela de ganho que é atribuída ao dito sinal de saída de áudio, como uma de duas ou mais funções de ganho, em que o modificador de sinais pode, por exemplo, ser configurado para gerar o dito sinal de saída de áudio de acordo com a dita função de janela de ganho, e em que, se o valor do argumento da dita função de janela de ganho é maior do que um limite de janela inferior e menor do que um limite de janela superior, a função de janela de ganho é configurada para retornar um valor de retorno de função de ganho que é maior do que qualquer valor de retorno de função de ganho retornado pela dita função de janela de ganho, se o valor de argumento de função de janela for menor do que o limite inferior, ou maior do que o limite superior.

[046] Em uma modalidade, a função de janela de ganho de cada um de dois ou mais sinais de saída de áudio tem um ou mais máximos globais, sendo um dos valores de argumento de função de ganho da dita função de janela de ganho, em que, para cada um de um ou mais máximos globais da dita função de janela de ganho, não existe nenhum outro valor de argumento de função de ganho para

o qual a dita função de janela de ganho retorna um valor de retorno de função de ganho maior do que para o dito máximo global, e em que, para cada par de um primeiro sinal de saída de áudio e um segundo sinal de saída de áudio dos dois ou mais sinais de saída de áudio, pelo menos um de um ou mais máximos globais da função de janela de ganho do primeiro sinal de saída de áudio pode, por exemplo, ser igual a um de um ou mais máximos globais da função de janela de ganho do segundo sinal de saída de áudio.

[047] De acordo com uma modalidade, o módulo de computação de função de ganho pode, por exemplo, ser configurado para receber, adicionalmente, informações de orientação, indicando um desvio angular de uma direção de busca em relação à direção de chegada, e em que o módulo de computação de função de ganho pode, por exemplo, ser configurado para gerar a função de ganho de panning de cada um dos sinais de saída de áudio, dependendo das informações de orientação.

[048] Em uma modalidade, o módulo de computação de função de ganho pode, por exemplo, ser configurado a gerar a função de janela de ganho de cada um dos sinais de saída de áudio dependendo das informações de orientação.

[049] De acordo com uma modalidade, o módulo de computação de função de ganho pode, por exemplo, ser configurado para receber informações de zoom, em que as informações de zoom indicam um ângulo de abertura de uma câmara, e em que o módulo de computação de função de ganho pode, por exemplo, ser configurado para gerar a função de ganho de panning de cada um dos sinais de saída de áudio, dependendo das informações de zoom.

[050] Em uma modalidade, o módulo de computação de função de ganho pode, por exemplo, ser configurado para gerar a função de janela de ganho de cada um dos sinais de saída de áudio, dependendo das informações de zoom.

[051] De acordo com uma modalidade, o módulo de computação de função de ganho pode, por exemplo, ser configurado para receber, adicionalmente, um parâmetro de calibração para alinhar uma imagem visual e uma imagem acústica, e em que o módulo de computação de função de ganho pode, por exemplo, ser configurado para gerar a função de ganho de panning de cada um dos sinais de saída de áudio, dependendo do parâmetro de calibração.

[052] Em uma modalidade, o módulo de computação de função de ganho

pode, por exemplo, ser configurado para gerar a função de janela de ganho de cada um dos sinais de saída de áudio, dependendo do parâmetro de calibração.

[053] Um sistema de acordo com uma das reivindicações anteriores, o módulo de computação de função de ganho pode, por exemplo, ser configurado para receber informações sobre uma imagem visual, e o módulo de computação de função de ganho pode, por exemplo, ser configurado para gerar, dependendo das informações sobre uma imagem visual, uma função de desfocagem que retorna ganhos complexos para realizar a dispersão perceptiva de uma fonte de som.

[054] Além disso, um aparelho para gerar um ou mais sinais de saída de áudio é fornecido. O aparelho compreende um processador de sinais e uma interface de saída. O processador de sinais é configurado para receber um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de áudio originais, em que o processador de sinais é configurado para receber um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de áudio originais, e em que o processador de sinais é configurado para receber informação de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio. Além disso, o processador de sinais é configurado para gerar um ou mais sinais difusos processados dependendo do sinal de componente de desarme. Para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, o processador de sinais é configurado para determinar, dependendo da direção de chegada, um ganho direto, o processador de sinais é configurado para aplicar dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e o processador de sinais é configurado para combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio. A interface de saída é configurada para produzir o um ou mais sinais de saída de áudio. O processador de sinais compreende um módulo de computação de função de ganho para calcular uma ou mais funções de ganho, em que cada função de ganho de uma ou mais funções de ganho, compreende uma pluralidade de valores de argumento de função de ganho, em que um valor de retorno de função de ganho é atribuído a cada um dos ditos valores de

argumento de função de ganho, em que, quando a dita função de ganho recebe um dos ditos valores de argumento de função de ganho, em que a dita função de ganho é configurada para retornar o valor de retorno de função de ganho que está sendo atribuído a um dos ditos valores de argumento de função de ganho. Além disso, o processador de sinais compreende, adicionalmente, um modificador de sinais para selecionar, dependendo da direção de chegada, um valor de argumento dependente de direção a partir dos valores de argumento de função de ganho de uma função de ganho de uma ou mais funções de ganho, para obter o valor de retorno de função de ganho sendo atribuído ao dito valor de argumento dependente de direção a partir da dita função de ganho, e para determinar o valor de ganho de pelo menos um de um ou mais sinais de saída de áudio, dependendo do dito valor de retorno de função de ganho obtido a partir da dita função de ganho.

[055] Além disso, um método para gerar um ou mais sinais de saída de áudio é fornecido. O método compreende:

[056] - Receber dois ou mais sinais de entrada de áudio.

[057] - Gerar um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de entrada de áudio.

[058] - Gerar um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de entrada de áudio.

[059] - Receber informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio.

[060] - Gerar um ou mais sinais difusos processados, dependendo do sinal de componente de desarme.

[061] - Para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, determinar, dependendo da direção de chegada, um ganho direto, aplicar o dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio. E:

[062] - Produzir o um ou mais sinais de saída de áudio.

[063] A geração de um ou mais sinais de saída de áudio compreende o cálculo de uma ou mais funções de ganho, em que cada função de ganho de

uma ou mais funções de ganho, compreende uma pluralidade de valores de argumento de função de ganho, em que um valor de retorno de função de ganho é atribuído para cada um dos ditos valores de argumento de função de ganho, em que, quando a dita função de ganho recebe um dos ditos valores de argumento de função de ganho, a dita função de ganho é configurada para retornar o valor de retorno de função de ganho que está sendo atribuído a um dos ditos valores de argumento de função de ganho. Além disso, a geração de um ou mais sinais de saída de áudio compreende selecionar, dependendo da direção de chegada, um valor de argumento dependente de direção a partir dos valores de argumento de função de ganho de uma função de ganho de uma ou mais funções de ganho, para obter o valor de retorno de função de ganho sendo atribuído ao dito valor de argumento dependente de direção, a partir da dita função de ganho, e para determinar o valor de ganho de pelo menos um de um ou mais sinais de saída de áudio, dependendo do dito valor de retorno de função de ganho obtido a partir da dita função de ganho.

[064] Além disso, um método para gerar um ou mais sinais de saída de áudio é fornecido. O método compreende:

[065] - Receber um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de áudio originais.

[066] - Receber um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de áudio originais.

[067] - Receber informações de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio.

[068] - Gerar um ou mais sinais difusos processados, dependendo do sinal de componente de desarme.

[069] - Para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, determinar, dependendo da direção de chegada, um ganho direto, aplicar o dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio. E:

[070] - Produzir o um ou mais sinais de saída de áudio.

[071] A geração dos sinais de um ou mais de áudio de saída compreende

o cálculo de uma ou mais funções de ganho, em que cada função de ganho de uma ou mais funções de ganho, compreende uma pluralidade de valores de argumento de função de ganho, em que um valor de retorno de função de ganho é atribuído a cada um dos ditos valores de argumento de função de ganho, em que, quando a dita função de ganho recebe um dos ditos valores de argumento de função de ganho, a dita função de ganho é configurada para retornar o valor de retorno de função de ganho que está sendo atribuído a um dos ditos valores de argumento de função de ganho. Além disso, a geração de um ou mais sinais de saída de áudio compreende selecionar, dependendo da direção de chegada, um valor de argumento dependente de direção a partir dos valores de argumento de função de ganho de uma função de ganho de uma ou mais funções de ganho, para obter o valor de retorno de função de ganho sendo atribuído ao dito valor de argumento dependente de direção a partir da dita função de ganho, e para determinar o valor de ganho de pelo menos um de um ou mais sinais de saída de áudio, dependendo do dito valor de retorno de função de ganho obtido a partir da dita função de ganho.

[072] Além disso, os programas de computador são fornecidos, em que cada um dos programas de computador é configurado para implementar um dos métodos acima descritos quando estiver sendo executado em um computador ou processador de sinais, de modo que cada um dos métodos acima descrito é executado por um dos programas de computador.

[073] A seguir, as modalidades da presente invenção são descritas em mais detalhe com referência às figuras, nas quais:

[074] A Fig. 1a ilustra um sistema de acordo com uma modalidade,

[075] A Fig. 1b ilustra um aparelho de acordo com uma modalidade,

[076] A Fig. 1C ilustra um sistema de acordo com outra modalidade,

[077] A Fig. 1D ilustra um aparelho de acordo com outra modalidade,

[078] A Fig. 2 mostra um sistema de acordo com outra modalidade,

[079] A Fig. 3 representa os módulos para a decomposição direta/difusa e para parâmetro de estimativa de um sistema de acordo com a uma modalidade,

[080] A Fig. 4 mostra uma primeira geometria para a reprodução de cena acústica com zoom acústico de acordo com uma modalidade, na qual uma fonte de som está localizada em um plano focal,

[081] A Fig. 5 ilustra funções panning para reprodução de cena consistente e para zoom acústico,

[082] A Fig. 6 representa funções panning adicionais para reprodução de cena consistente e para zoom acústico de acordo com as modalidades,

[083] A Fig. 7 ilustra funções de janela de ganho exemplificadoras para várias situações de acordo com modalidades,

[084] A Fig. 8 mostra uma função de ganho difuso de acordo com uma modalidade,

[085] A Fig. 9 representa uma segunda geometria para a reprodução de cena acústica com zoom acústico de acordo com uma modalidade, em que uma fonte de som não está localizada em um plano focal,

[086] A Fig. 10 ilustra as funções para explicar a desfocagem de som direto, e

[087] A Fig. 11 visualiza aparelhos auditivos de acordo com modalidades.

[088] A Fig. 1a ilustra um sistema para gerar um ou mais sinais de saída de áudio. O sistema compreende um módulo de decomposição 101, um processador de sinais 105, e uma interface de saída 106.

[089] O módulo de decomposição 101 é configurado para gerar um sinal de componente direto $X_{dir}(k, n)$, que compreende componentes de sinal direto de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$. Além disso, o módulo de decomposição 101 é configurado para gerar um sinal de componente difuso $X_{diff}(k, n)$, que compreende componentes de sinal difuso de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$.

[090] O processador de sinais 105 é configurado para receber o sinal de componente direto $X_{dir}(k, n)$, o sinal de componente difuso $X_{diff}(k, n)$ e as informações de direção, as ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto dos dois ou mais de entrada de áudio sinaliza $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$.

[091] Além disso, o processador de sinais 105 é configurado para gerar um ou mais sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, …, $Y_{diff,v}(k, n)$, dependendo do sinal de componente de desarme $X_{diff}(k, n)$.

[092] Para cada sinal de saída de áudio $Y_i(k, n)$, de um ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$, o processador de sinais 105 é

configurado para determinar, dependendo da direção de chegada, um ganho direto $G_i(k, n)$, o processador de sinais 105 é configurado para aplicar dito ganho direto $G_i(k, n)$ sobre o sinal de componente direto $X_{dir}(k, n)$ para obter um sinal direto processado $Y_{dir,i}(k, n)$, e o processador de sinais 105 é configurado para combinar o dito sinal direto processado $Y_{dir,i}(k, n)$ e um $Y_{diff,i}(k, n)$, de um ou sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, …, $Y_{diff,v}(k, n)$ para gerar o dito sinal de saída de áudio $Y_i(k, n)$.

[093] A interface de saída 106 é configurada para produzir um ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$.

[094] Tal como dito, as informações de direção dependem de uma direção de chegada $\varphi(k, n)$ dos componentes de sinal direto de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$.. Por exemplo, a direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$ podem, por exemplo, ser por si só as informações de direção. Ou, por exemplo, as informações de direção, podem, por exemplo, ser a direção de propagação dos componentes de sinal direto de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$. Embora a direção de chegada aponte de uma matriz de microfones de recepção para uma fonte de som, a direção de propagação aponta da fonte de som para a matriz de microfones de recepção. Desse modo, a direção de propagação aponta exatamente na direção oposta da direção de chegada e, como consequência, depende da direção de chegada.

[095] Para gerar um $Y_i(k, n)$ de um ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$ o processador de sinais 105

[096] - determinar, dependendo da direção de chegada, um ganho direto $G_i(k, n)$,

[097] - aplicar o dito ganho direto $G_i(k, n)$ no sinal de componente direto $X_{dir}(k, n)$ para obter um sinal direto processado $Y_{dir,i}(k, n)$, e

[098] - combinar o dito sinal direto processado $Y_{dir,i}(k, n)$ e um $Y_{diff,i}(k, n)$ de um ou mais sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, …, $Y_{diff,v}(k, n)$ para gerar o dito sinal de saída de áudio $Y_i(k, n)$

[099] Isto é feito para cada um de um ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$ que devem ser gerados $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$. O

processador de sinais pode, por exemplo, ser configurado para gerar um, dois, três ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, ..., $Y_v(k, n)$.

[100] No que diz respeito a um ou mais sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, ..., $Y_{diff,v}(k, n)$, de acordo com uma modalidade, o processador de sinais 105 pode, por exemplo, ser configurado para gerar um ou mais sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, ..., $Y_{diff,v}(k, n)$ através da aplicação de um ganho difuso $Q(k, n)$ no sinal de componente difuso $X_{diff}(k, n)$.

[101] O módulo de decomposição 101 é configurado, por exemplo, para gerar o sinal de componente direto $X_{dir}(k, n)$, compreendendo os componentes de sinal direto de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, ... $x_p(k, n)$, e o sinal de componente difuso $X_{diff}(k, n)$, que compreende componentes de sinal difuso de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, ... $x_p(k, n)$ pela decomposição de um ou mais sinais de entrada de áudio para o sinal de componente direto e para o sinal de componente difuso.

[102] Em uma modalidade particular, o processador de sinais 105 pode, por exemplo, ser configurado para gerar dois ou mais canais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, ..., $Y_v(k, n)$. O processador de sinais 105 pode, por exemplo, ser configurado para aplicar o ganho difuso $Q(k, n)$ no sinal de componente difuso $X_{diff}(k, n)$ para obter um sinal difuso intermediário. Além disso, o processador de sinais 105 pode, por exemplo, ser configurado para gerar um ou mais sinais descorrelacionados a partir do sinal difuso intermediário através da condução de descorrelação, em que o um ou mais sinais descorrelacionados formam um ou mais sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, ..., $Y_{diff,v}(k, n)$, ou em que o sinal difuso intermediário e o um ou mais sinais descorrelacionados formam um ou mais sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, ..., $Y_{diff,v}(k, n)$.

[103] Por exemplo, o número de sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, ..., $Y_{diff,v}(k, n)$ e o número de sinais de saída de áudio pode, por exemplo, ser igual $Y_1(k, n)$, $Y_2(k, n)$, ..., $Y_v(k, n)$.

[104] A geração de um ou mais sinais descorrelacionados do sinal difuso intermediário pode, por exemplo, ser conduzida por aplicação de retardos no sinal difuso intermediário, ou, por exemplo, por convolução do sinal difuso intermediário com uma rajada de ruído, ou, por exemplo, por convolução do sinal difuso intermediário com uma resposta ao impulso, etc. Qualquer outra técnica

de descorrelação do estado da técnica pode, por exemplo, alternativamente ou adicionalmente ser aplicada.

[105] Para a obtenção de sinais de saída de áudio v $Y_1(k, n)$, $Y_2(k, n)$, ..., $Y_v(k, n)$, as determinações de v dos ganhos diretos de v $G_1(k, n)$, $G_2(k, n)$, ..., $G_v(k, n)$ e aplicações de v do respectivo ganho sobre um ou mais sinais de componente diretos $X_{dir}(k, n)$ podem, por exemplo, ser usadas para obter os sinais de saída de áudio v $Y_1(k, n)$, $Y_2(k, n)$, ..., $Y_v(k, n)$.

[106] Apenas um único sinal de componente difuso $X_{diff}(k, n)$, apenas uma única determinação de um ganho difuso $Q(k, n)$ e apenas uma única aplicação de ganho difuso $Q(k, n)$ sobre o sinal de componente difuso $X_{diff}(k, n)$ podem, por exemplo, ser necessárias, para obter os sinais de saída de áudio v $Y_1(k, n)$, $Y_2(k, n)$, ..., $Y_v(k, n)$. Para obter a descorrelação, técnicas de descorrelação podem ser aplicadas apenas após o ganho difuso já ter sido aplicado no sinal de componente difuso.

[107] De acordo com a modalidade da Fig. 1a, o mesmo sinal difuso processado $Y_{diff}(k, n)$ é então combinado com um ($Y_{dir,i}(k, n)$) correspondente dos sinais diretos processados para obter o um ($Y_i(k, n)$) correspondente dos sinais de saída de áudio.

[108] A modalidade da Fig. 1a leva em consideração a direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, ... $x_p(k, n)$. Desse modo, os sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, ..., $Y_v(k, n)$ podem ser gerados pelo ajuste de flexibilidade dos sinais de componente diretos $X_{dir}(k, n)$ e os sinais de componente difusos $X_{diff}(k, n)$ dependendo da direção de chegada. As possibilidades de adaptação avançadas são alcançadas.

[109] De acordo com as modalidades, os sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, ..., $Y_v(k, n)$ podem, por exemplo, ser determinados para cada binário de tempo-frequência (k, n) de um domínio de tempo -frequência.

[110] De acordo com uma modalidade, o módulo de decomposição 101 pode, por exemplo, ser configurado para receber dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, ... $x_p(k, n)$. Em uma outra modalidade, o módulo de decomposição 101 pode, por exemplo, ser configurado para receber três ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, ... $x_p(k, n)$. O módulo de

decomposição 101 pode, por exemplo, ser configurado para decompor os dois ou mais (ou três ou mais sinais de entrada de áudio) $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$ para o sinal de componente difuso $X_{diff}(k, n)$, que não é um sinal multicanais, e para o um ou mais sinais de componente diretos $X_{dir}(k, n)$. O fato de um sinal de áudio não ser um sinal multicanal significa que o sinal de áudio por si só não compreende mais do que um canal de áudio. Desse modo, as informações de áudio da pluralidade de sinais de entrada de áudio são transmitidas dentro dos dois sinais de componente ($X_{dir}(k, n)$, $X_{diff}(k, n)$) (e, possivelmente, em informações laterais adicionais), que permite a transmissão eficiente.

[111] O processador de sinais 105, pode, por exemplo, ser configurado para gerar cada sinal de saída de áudio $Y_i(k, n)$ de dois ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$ através da determinação do ganho direto $G_i(k, n)$ para o dito sinal de saída de áudio $Y_i(k, n)$, através da aplicação do dito ganho direto $G_i(k, n)$, sobre o um ou mais sinais de componente diretos $X_{dir}(k, n)$ para obter o sinal direto processado $Y_{dir,i}(k, n)$ para o dito sinal de saída de áudio $Y_i(k, n)$, e, através da combinação do dito sinal direto processado $Y_{dir,i}(k, n)$ para o dito sinal de saída de áudio $Y_i(k, n)$ e o sinal difuso processado $Y_{diff}(k, n)$ para gerar o dito sinal de saída de áudio $Y_i(k, n)$. A interface de saída 106 é configurada para produzir os dois ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$. A geração de dois ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$ através da determinação de apenas um único sinal difuso processado $Y_{diff}(k, n)$ é particularmente vantajosa.

[112] A Fig. 1b ilustra um aparelho para gerar um ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$ de acordo com uma modalidade. O aparelho implementa o chamado lado "da extremidade distante" do sistema da Fig. 1-A.

[113] O aparelho da Fig. 1b compreende um processador de sinais 105, e uma interface de saída 106.

[114] O processador de sinais 105 é configurado para receber um sinal de componente direto $X_{dir}(k, n)$, compreendendo componentes de sinal direto, de dois ou mais sinais de áudio originais $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$ (por exemplo, os sinais de entrada de áudio da Fig. 1a). Além disso, o processador de sinais 105 é configurado para receber um sinal de componente difuso $X_{diff}(k, n)$, que

compreende componentes de sinal difuso de dois ou mais sinais de áudio originais $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$. Além disso, o processador de sinais 105 é configurado para receber as informações de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio.

[115] O processador de sinais 105 é configurado para gerar um ou mais sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, …, $Y_{diff,v}(k, n)$ dependendo do sinal de componente de desarme $X_{diff}(k, n)$.

[116] Para cada sinal de saída de áudio $Y_i(k, n)$ de um ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$, o processador de sinais 105 é configurado para determinar, dependendo da direção de chegada, um ganho direto $G_i(k, n)$, o processador de sinais 105 é configurado para aplicar dito ganho direto $G_i(k, n)$ no sinal de componente direto $X_{dir}(k, n)$ para obter um sinal direto processado $Y_{dir,i}(k, n)$, e o processador de sinais 105 é configurado para combinar o dito sinal direto processado $Y_{dir,i}(k, n)$ e um $Y_{diff,i}(k, n)$ de um ou sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, …, $Y_{diff,v}(k, n)$ para gerar o dito sinal de saída de áudio $Y_i(k, n)$.

[117] A interface de saída 106 é configurada para produzir um ou mais sinais de saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$.

[118] Todas as configurações do processador de sinais 105 descritas com referência ao sistema a seguir, também podem ser implementadas num aparelho de acordo com a Fig. 1b. Isto diz respeito, em particular, às várias configurações de modificador de sinais 103 e módulo de computação de função de ganho 104, que são descritas abaixo. O mesmo se aplica para os vários exemplos de aplicação dos conceitos descritos a seguir.

[119] A Fig. 1c ilustra um sistema de acordo com outra modalidade. Na Fig. 1c, o gerador de sinal 105 da Fig. 1a compreende, adicionalmente, um módulo de computação de função de ganho 104 para calcular uma ou mais funções de ganho, em que cada função de ganho de uma ou mais funções de ganho, compreende uma pluralidade de valores de argumento de função de ganho, em que um valor de retorno de função de ganho é atribuído a cada um dos ditos valores de argumento de função de ganho, em que, quando a dita função de ganho recebe um dos ditos valores de argumento de função de ganho, em que

a dita função de ganho é configurada para retornar o valor de retorno de função de ganho sendo atribuído a um dos ditos valores de argumento de função de ganho.

[120] Além disso, o processador de sinais 105 compreende, adicionalmente, um modificador de sinais 103 para selecionar, dependendo da direção de chegada, um valor de argumento dependente de direção, a partir dos valores de argumento de função de ganho de uma função de ganho de uma ou mais funções de ganho, para a obtenção do valor de retorno de função de ganho sendo atribuído ao dito valor de argumento dependente de direção a partir da dita função de ganho, e para determinação do valor de ganho de pelo menos um de um ou mais sinais de saída de áudio, dependendo do dito valor de retorno de função de ganho obtido a partir da dita função de ganho.

[121] A Fig. 1d ilustra um sistema de acordo com uma outra modalidade. Na Fig. 1d, o gerador de sinal 105 da Fig. 1b compreende, adicionalmente, um módulo de computação de função de ganho 104 para calcular uma ou mais funções de ganho, em que cada função de ganho de uma ou mais funções de ganho, compreende uma pluralidade de valores de argumento de função de ganho, em que um valor de retorno de função de ganho é atribuído a cada um dos ditos valores de argumento de função de ganho, em que, quando a dita função de ganho recebe um dos ditos valores de argumento de função de ganho, em que a dita função de ganho é configurada para retornar o valor de retorno de função de ganho sendo atribuído a um dos ditos valores de argumento de função de ganho.

[122] Além disso, o processador de sinais 105 compreende, adicionalmente, um modificador de sinais 103 para selecionar, dependendo da direção de chegada, um valor de argumento dependente de direção a partir dos valores de argumento de função de ganho de uma função de ganho de uma ou mais funções de ganho, para a obtenção de valor de retorno de função de ganho sendo atribuído ao dito valor de argumento dependente de direção a partir da dita função de ganho, e para determinação do valor de ganho de pelo menos um de um ou mais sinais de saída de áudio, dependendo do dito valor de retorno de função de ganho obtido a partir da dita função de ganho.

[123] As modalidades fornecem a gravação e reprodução do som espacial

tal que a imagem acústica é consistente com uma imagem espacial desejada, que é determinada, por exemplo, por um vídeo que está complementando o áudio no lado de extremidade distante. Algumas modalidades são baseadas em gravações com uma matriz de microfones localizada no lado de extremidade próxima reverberante. As modalidades fornecem, por exemplo, um zoom acústico que é consistente com o zoom visual de uma câmara. Por exemplo, quando se aumenta o zoom, o som direto dos alto-falantes é reproduzido a partir da direção em que os alto-falantes estariam localizados na imagem visual ampliada, de tal modo que as imagens visuais e acústicas são alinhadas. Se os alto-falantes estão localizados fora da imagem visual (ou fora de uma região espacial desejada) após o aumento do zoom, o som direto destes alto-falantes pode ser atenuado, uma vez que estes alto-falantes não são mais visíveis, ou, por exemplo, já que o som direto destes alto-falantes não é desejado. Além disso, a razão de direto- para -reverberação pode, por exemplo, ser aumentada quando se aumenta o zoom para imitar o ângulo de abertura menor da câmera visual.

[124] As modalidades são baseadas no conceito de separar os sinais de microfone gravados no som direto das fontes de som e o som difuso, por exemplo, som de reverberação, através da aplicação de dois filtros multicanais recentemente no lado de extremidade próxima. Estes filtros multicanais podem, por exemplo, ser baseados em informações paramétricas do campo de som, tais como o DOA de som direto. Em algumas modalidades, o som direto e som difuso separados podem, por exemplo, ser transmitidos para o lado de extremidade distante em conjunto com as informações paramétricas.

[125] Por exemplo, no lado de extremidade distante, ponderações específicas podem, por exemplo, ser aplicadas ao som direto e som difuso extraídos, que ajustam a imagem acústica reproduzida de tal modo que os sinais de saída de áudio resultantes são consistentes com uma imagem espacial desejada. Estas ponderações modelam, por exemplo, o efeito de zoom acústico e dependem, por exemplo, da direção de chegada (DOA) do som direto e, por exemplo, de um fator de zoom e/ou de uma direção de busca de uma câmara. Os sinais de saída de áudio finais podem, por exemplo, em seguida, ser obtidos pela soma do som direto e do som difuso ponderados.

[126] Os conceitos fornecidos realizam um uso eficiente no cenário de gravação de vídeo acima mencionado com dispositivos de consumo ou em um cenário de teleconferência: Por exemplo, no cenário de gravação de vídeo, pode-se, por exemplo, ser suficiente armazenar ou transmitir o som direto e som difuso extraídos (em vez de todos os sinais de microfone), enquanto continuam sendo capazes de controlar a imagem espacial recriada.

[127] Isso significa que, se, por exemplo, um visual zoom é aplicado em uma etapa de pós-processamento (zoom digital), a imagem acústica pode ainda ser modificada em conformidade, sem a necessidade de armazenar e acessar os sinais de microfone originais. No cenário de teleconferência, os conceitos propostos também podem ser usados de forma eficiente, uma vez que a extração de som direto e difuso pode ser realizada no lado de extremidade próxima enquanto continua sendo capaz de controlar a reprodução de som espacial (por exemplo, alterando a configuração do alto-falante) no lado de extremidade distante e para alinhar a imagem acústica e visual. Portanto, é apenas necessário transmitir alguns sinais de áudio e as DOAs estimadas como informações de lado, embora a complexidade computacional no lado de extremidade distante seja baixa.

[128] A Fig. 2 ilustra um sistema de acordo com uma modalidade. O lado de extremidade próxima compreende os módulos 101 e 102. O lado de extremidade distante compreende os módulos 105 e 106. O módulo 105 em si compreende os módulos 103 e 104. Quando se faz referência a um lado de extremidade próxima e a um lado de extremidade distante, entende-se que, em algumas modalidades, um primeiro aparelho pode implementar o lado de extremidade próxima (por exemplo, compreendendo os módulos 101 e 102), e um segundo aparelho pode implementar o lado de extremidade distante (por exemplo, compreendendo os módulos 103 e 104), enquanto que em outras modalidades, um único aparelho implementa o lado de extremidade próxima, bem como o lado de extremidade distante, em que um único tal aparelho, por exemplo, compreende os módulos 101, 102, 103 e 104.

[129] Em particular, a Fig. 2 ilustra um sistema de acordo com uma modalidade que compreende um módulo de decomposição 101, um módulo de estimativa de parâmetro 102, um processador de sinais 105, e uma interface de

saída 106. Na Fig. 2, o processador de sinais 105 compreende um módulo de computação de função ganho 104 e um modificador de sinais 103. O processador de sinais 105 e a interface de saída 106 podem, por exemplo, compreender um aparelho tal como ilustrado pela Fig. 1b.

[130] Na Fig. 2, *inter alia*, o módulo de estimativa de parâmetro 102 pode, por exemplo, ser configurado para receber os dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$. Além disso, o módulo de estimativa de parâmetro 102 pode, por exemplo, ser configurado para estimar a direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$ dependendo de dois ou mais sinais de entrada de áudio. O processador de sinais 105 pode, por exemplo, ser configurado para receber as informações de direção de chegada compreendendo a direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio do módulo de estimativa de parâmetro 102.

[131] A entrada do sistema da Fig. 2 consiste em *M* sinais de microfone $X_{1...M}(k, n)$ no domínio do tempo-frequência (k é índice de frequência, n é índice de tempo). Pode, por exemplo, ser assumido que o campo de som, que é capturado pelos microfones, consiste, para cada *(k, n)* em uma propagação de onda plana em um campo difuso isotrópico. A onda plana modela o som direto das fontes de som (por exemplo, alto-falantes), enquanto o som difuso modela a reverberação.

[132] De acordo com um tal modelo deste tipo, o *m*-ésimo sinal de microfone pode ser escrito como

$$X_m(k,n) = X_{\mathrm{dir},m}(k,n) + X_{\mathrm{diff},m}(k,n) + X_{\mathrm{n},m}(k,n), \quad (1)$$

[133] onde $X_{dir,m}(k, n)$ é o som direto medido (onda plana), $X_{diff,m}(k, n)$ é o som difuso medido, e $X_{n,m}(k, n)$ é um componente de ruído (por exemplo, um autorruído de microfone).

[134] No módulo de decomposição 101 da Fig. 2 (decomposição direta/difusa), o som direto $X_{dir}(k, n)$ e som difuso $X_{diff}(k, n)$ são extraídos a partir dos sinais do microfone. Para este efeito, por exemplo, os filtros multicanais informados como descritos a seguir podem ser usados. Para a decomposição difusa/direta, as informações paramétricas específicas sobre o campo sonoro

podem, por exemplo, ser usadas, por exemplo, a DOA do som direto $\varphi(k, n)$. Esta informação paramétrica pode, por exemplo, ser calculada a partir dos sinais de microfone no módulo de estimativa de parâmetro 102. Além de DOA $\varphi(k, n)$ do som direto, em algumas modalidades, umas informações de distância $r(k, n)$ pode, por exemplo, ser estimada. Esta informação da distância pode, por exemplo, descrever a distância entre a matriz de microfone e a fonte de som, que está emitindo uma onda plana. Para a estimativa de parâmetros, estimadores de distância e/ou estimadores de DOA do estado da técnica, podem, por exemplo, ser usados. Os estimadores correspondentes podem, por exemplo, ser descritos abaixo.

[135] O som direto extraído $X_{dir}(k, n)$, o som difuso extraído $X_{diff}(k, n)$, e as informações paramétricas estimadas do som direto, por exemplo, DOA $\varphi(k, n)$ e/ou a distância $r(k, n)$, podem, por exemplo, em seguida, ser armazenadas, transmitidas para o lado de extremidade distante, ou imediatamente a ser usadas para gerar o som espacial com a imagem espacial desejada, por exemplo, para criar o efeito de zoom acústico.

[136] A imagem acústica desejada, por exemplo, um efeito de zoom acústico, é gerada no modificador de sinais 103 usando o som direto extraído $X_{dir}(k, n)$, o som difuso extraído $X_{diff}(k, n)$, e as informações paramétricas estimadas $\varphi(k, n)$ e/ou $r(k, n)$.

[137] O modificador de sinais 103 pode, por exemplo, calcular um ou mais sinais de saída $Y_i(k, n)$ no domínio do tempo-frequência que recriar a imagem acústica tal que ela seja consistente com a imagem espacial desejada. Por exemplo, os sinais de saída $Y_i(k, n)$ imitam o efeito de zoom acústico. Estes sinais podem ser finalmente transformados de volta para o domínio do tempo e reproduzido, por exemplo, por alto-falantes ou fones de ouvido. O *i*-ésimo sinal de saída $Y_i(k, n)$ é calculado como uma soma ponderada do som direto $X_{dir}(k, n)$ e do som difuso $X_{diff}(k, n)$ extraídos, por exemplo,

$$Y_i(k,n) = G_i(k,n)X_{\mathrm{dir}}(k,n) + f_i\Big\{\underbrace{QX_{\mathrm{diff}}(k,n)}_{Y_{\mathrm{diff}}(k,n)}\Big\} \tag{2a}$$

$$= Y_{\mathrm{dir},i}(k,n) + Y_{\mathrm{diff},i}(k,n). \tag{2b}$$

[138] Nas fórmulas (2a) e (2b), as ponderações $G_i(k, n)$ e Q são parâmetros

que são usados para criar a imagem acústica desejada, por exemplo, o efeito de zoom acústico. Por exemplo, quando se aumenta o zoom, o parâmetro Q pode ser reduzido de modo que o som difuso reproduzido é atenuado.

[139] Além disso, com as ponderações $G_i(k, n)$ é possível o controle de qual direção o som direto é reproduzido de modo que a imagem visual e acústica é alinhada. Além disso, um efeito de desfocagem acústico pode ser alinhado ao som direto.

[140] Em algumas modalidades, as ponderações $G_i(k, n)$ e Q podem, por exemplo, ser determinadas em unidades de seleção de ganho 201 e 202. Estas unidades podem, por exemplo, selecionar as ponderações $G_i(k, n)$ e Q apropriados a partir de duas funções de ganho, denotadas por g_i e q, dependendo da informação paramétrica estimada $\varphi(k, n)$ e $r(k, n)$. Expresso matematicamente,

$$G_i(k,n) = g_i(\varphi, r), \quad (3a)$$

$$Q(k,n) = q(r). \quad (3b)$$

[141] Em algumas modalidades, as funções de ganho g_i e q podem depender da aplicação e podem, por exemplo, ser geradas no módulo de computação de função de ganho 104. As funções de ganho descrevem que as ponderações $G_i(k, n)$ e Q devem ser usadas em (2a) para uma determinada informação paramétrica $\varphi(k, n)$ e/ou $r(k, n)$ de tal modo que as imagens espaciais consistentes desejadas são obtidas.

[142] Por exemplo, quando se aumenta o zoom com a câmara visual, as funções de ganho são ajustadas de modo que o som é reproduzido a partir das direções em que as fontes são visíveis no vídeo. As ponderações $G_i(k, n)$ e Q e as funções de ganho subjacentes g_i e q são adicionalmente descritas abaixo. Deve-se notar que as ponderações $G_i(k, n)$ e Q e as funções de ganho subjacentes g_i e q podem, por exemplo, ser de valor complexo. O cálculo das funções de ganho requer informações como o fator de zoom, a largura da imagem visual, a direção de busca desejada e a configuração alto-falante.

[143] Em outras modalidades, as ponderações são $G_i(k, n)$ e Q são diretamente calculadas dentro do modificador de sinais 103, em vez de no início do cálculo das funções de ganho no módulo 104 e, em seguida, seleciona-se as

ponderações $G_i(k, n)$ e Q a partir das funções de ganho calculadas nas unidades de seleção de ganho 201 e 202.

[144] De acordo com modalidades, mais de uma onda plana por tempo-frequência pode, por exemplo, ser especificamente processada. Por exemplo, duas ou mais ondas planas na mesma banda de frequência de duas direções diferentes podem, por exemplo, chegar a ser gravadas por uma matriz de microfones no mesmo ponto no tempo. Estas duas ondas planas podem, cada uma, ter uma direção diferente de chegada. Em tais cenários, os componentes de sinal direto de duas ou mais ondas planas e a sua direção de chegadas podem, por exemplo, ser considerados separadamente.

[145] De acordo com modalidades, o sinal de componente direto $X_{dir1}(k, n)$ e um ou mais sinais de componente diretos adicionais $X_{dir2}(k, n)$, …, $X_{dir\,q}(k, n)$ podem, por exemplo, formar um grupo de dois ou mais sinais de componente diretos $X_{dir1}(k, n)$, $X_{dir2}(k, n)$, …, $X_{dir\,q}(k, n)$, em que o módulo de decomposição 101 pode, por exemplo, ser configurado para gerar um ou mais sinais de componente diretos $X_{dir2}(k, n)$, …, $X_{dir\,q}(k, n)$ compreendendo componentes de sinal direto adicionais de dois ou mais sinais de entrada de áudio $x_1(k, n)$, $x_2(k, n)$, … $x_p(k, n)$.

[146] A direção de chegada e uma ou mais direções de chegadas adicionais formam um grupo de duas ou mais direções de chegadas, em que cada direção de chegada do grupo de duas ou mais direções de chegadas é atribuída a exatamente um sinal de componente direto $X_{dir\,j}(k, n)$ do grupo de dois ou mais sinais de componente diretos $X_{dir1}(k, n)$, $X_{dir2}(k, n)$, …, $X_{dir\,q,m}(k, n)$, em que o número de sinais de componente diretos de dois ou mais sinais de componente diretos e o número da direção de chegadas das ditas duas direções de chegadas é igual.

[147] O processador de sinais 105 pode, por exemplo, ser configurado para receber o grupo de dois ou mais sinais de componente diretos $X_{dir1}(k, n)$, $X_{dir2}(k, n)$, …, $X_{dir\,q}(k, n)$, e o grupo de duas ou mais direções de chegadas.

[148] Para cada sinal de saída de áudio $Y_i(k, n)$ de um ou mais sinais saída de áudio $Y_1(k, n)$, $Y_2(k, n)$, …, $Y_v(k, n)$,

[149] - O processador de sinais 105 pode, por exemplo, ser configurado para determinar, para cada sinal de componente direto $X_{dir\,j}(k, n)$ do grupo de

dois ou mais sinais de componente diretos $X_{dir1}(k, n)$, $X_{dir2}(k, n)$, ..., $X_{dir\,q}(k, n)$, um ganho direto $G_{j,i}(k, n)$ dependendo da direção de chegada do dito sinal de componente direto $X_{dir\,j}(k, n)$,

[150] - O processador de sinais 105 pode, por exemplo, ser configurado para gerar um grupo de dois ou mais sinais diretos processados $Y_{dir1,i}(k, n)$, $Y_{dir2,i}(k, n)$, ..., $Y_{dir\,q,i}(k, n)$ através da aplicação, para cada sinal de componente direto $X_{dir\,j}(k, n)$ do grupo de dois ou mais sinais de componente diretos $X_{dir1}(k, n)$, $X_{dir2}(k, n)$, ..., $X_{dir\,q}(k, n)$, o ganho direto $G_{j,i}(k, n)$ do dito sinal de componente direto $X_{dir\,j}(k, n)$ no dito sinal de componente direto $X_{dir\,j}(k, n)$. E:

[151] - O processador de sinais 105 pode, por exemplo, ser configurado para combinar um $Y_{diff,i}(k, n)$ de um ou mais sinais difusos processados $Y_{diff,1}(k, n)$, $Y_{diff,2}(k, n)$, ..., $Y_{diff,v}(k, n)$ e cada sinal processado $Y_{dir\,j,i}(k, n)$ do grupo de dois ou mais sinais processados $Y_{dir1,i}(k, n)$, $Y_{dir2,i}(k, n)$, ..., $Y_{dir\,q,i}(k, n)$ para gerar o dito sinal de saída de áudio $Y_i(k, n)$.

[152] Desse modo, se duas ou mais ondas planas são consideradas separadamente, o modelo de fórmula (1) torna-se:

$$X_m(k, n) = X_{dir1,m}(k, n) + X_{dir2,m}(k, n) + \ldots + X_{dir\,q,m}(k, n) + X_{diff,m}(k, n) + X_{n,m}(k, n)$$

[153] e as ponderações podem, por exemplo, ser calculadas de forma análoga às fórmulas (2a) e (2b) de acordo com:

$$Y_i(k, n) = G_{1,i}(k, n)\, X_{dir1}(k, n) + G_{2,i}(k, n)\, X_{dir2}(k, n) + \ldots + G_{q,i}(k, n)\, X_{dir\,q}(k, n) + Q\, X_{diff,m}(k, n)$$
$$= Y_{dir1,i}(k, n) + Y_{dir2,i}(k, n) + \ldots + Y_{dir\,q,i}(k, n) + Y_{diff,i}(k, n)$$

[154] É suficiente que apenas alguns sinais de componente diretos, sinal de componente difuso e informações de lado sejam transmitidos de um lado de extremidade próxima para um lado de extremidade distante. Em uma modalidade, o número de sinal(s) de componente direto do grupo de dois ou mais sinais de componente diretos $X_{dir1}(k, n)$, $X_{dir2}(k, n)$, ..., $X_{dir\,q}(k, n)$ e 1 é menor do que o número de sinais de áudio de entrada $x_1(k, n)$, $x_2(k, n)$, ... $x_p(k, n)$ sendo recebido pela interface de recepção 101. (Usando os índices: $q + 1 < p$) "e 1" representa o sinal de componente difuso $X_{diff}(k, n)$ que é necessário.

[155] Quando a seguir, explicações são fornecidas com respeito a uma única onda plana, para uma única direção de chegada e para um único sinal de componente direto, deve ser entendido que os conceitos explicados são

igualmente aplicáveis a mais do que uma onda plana, mais do que uma direção de chegada e mais do que um componente de sinal direto.

[156] A seguir, a extração de som direto e difuso é descrita. As realizações práticas do módulo de decomposição 101 da Fig. 2, que realiza a decomposição direta/difusa, são fornecidas.

[157] Nas modalidades, para realizar a reprodução de som espacial constante, a produção de dois filtros de Variância Mínima Constrangida Linearmente informados recentemente propostos (linearly constrained minimum variance, LCMV) descritos em [8] e [9] são combinados, o que permite uma extração multicanais precisa de som direto e som difuso com uma resposta arbitrária desejada assumindo um modelo de campo de som semelhante ao de DirAC (Directional Audio Coding - Codificação Direcional de Áudio). Uma forma específica de combinar estes filtros de acordo com uma modalidade é agora descrita a seguir:

[158] Primeiramente, é descrita a extração de som direto de acordo com uma modalidade.

[159] O som direto é extraído usando o filtro espacial informado recentemente proposto descrito em [8]. Este filtro é brevemente descrito a seguir e, em seguida, formulado de tal forma que ele pode ser usado nas modalidades de acordo com a Fig. 2.

[160] O sinal direto desejado estimado $\hat{Y}_{dir,i}(k,n)$ para o *i*-ésimo canal de alto-falante em (2b) e a Fig. 2 é calculado pela aplicação de um filtro multicanais linear para os sinais do microfone, por exemplo,

$$\widehat{Y}_{\mathrm{dir},i}(k,n) = \mathbf{w}^{\mathrm{H}}_{\mathrm{dir},i}(k,n)\mathbf{x}(k,n), \quad (4)$$

[161] em que o vetor $\mathbf{x}(k, n) = [X_1(k, n), \ldots, X_M(k, n)]^T$ compreende os sinais de microfone M e $\mathbf{w}_{dir,i}$, é um vetor ponderado de valor complexo. Aqui, as ponderações de filtros minimizam o som difuso e o ruído compreendido nos microfones durante a captura do som direto com o ganho desejado $G_i(k, n)$. Expressas matematicamente, as ponderações, podem, por exemplo, ser calculadas como

$$\mathbf{w}_{\mathrm{dir},i}(k,n) = \arg\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\mathbf{\Phi}_{\mathrm{u}}(k,n)\mathbf{w} \quad (5)$$

[162] sujeito à restrição linear

$$\mathbf{w}^{\mathrm{H}}\mathbf{a}(k,\varphi) = G_i(k,n). \quad (6)$$

[163] Aqui, **a**(*k*, *φ*) é o chamado vetor de propagação de matriz. O *m*-ésimo elemento deste vetor é a função de transferência relativa do som direto entre o *m*-ésimo microfone e um microfone de referência da matriz (sem perda de generalidade do primeiro microfone na posição, $\mathbf{d}_1$ é usado na descrição seguinte). Este vetor depende de DOA *φ*(*k, n*) do som direto.

[164] O vetor de propagação de matriz é, por exemplo, definido em [8]. Na fórmula (6) do documento de [8], o vetor de propagação de matriz é definido de acordo com

$$\mathbf{a}(k,\varphi_l) = [a_1(k,\varphi_l) \ \ldots \ a_M(k,\varphi_l)]^{\mathrm{T}}$$

[165] em que φ_l é um ângulo de azimute de uma direção de chegada de uma *l*-ésima onda plana. Desse modo, o vetor de propagação de matriz depende da direção de chegada. Se apenas uma onda plana existe ou é considerada, o índice *l* pode ser omitido.

[166] De acordo com a fórmula (6) de [8], o elemento *i*-ésimo a_i do vetor de propagação de matriz **a** descreve o desvio de fase de uma *l*-ésima onda plana a partir de um primeiro para um *i*-ésimo microfone é definido de acordo com

$$a_i(k,\varphi_l) = \exp\{\jmath\,\kappa\, r_i \sin\varphi_l(k,n)\}.$$

[167] Por exemplo, r_i é igual a uma distância entre o primeiro e o *i*-ésimo microfone, *κ* indica o número de onda da onda plana e $\jmath$ é o número imaginário.

[168] Mais informações sobre o vetor de uma propagação matriz **a** e os seus elementos a_i podem ser encontradas em [8] que é expressamente aqui incorporado por referência.

[169] A matriz *M* × *M* $\mathbf{\Phi}_{\mathrm{u}}(k, n)$ em (5) é a matriz de densidade de potência espectral (power spectral density, PSD), do ruído e do som difuso, que pode ser determinada tal como explicado em [8]. A solução para (5) é dada por

$$\mathbf{w}_{\mathrm{dir},i}(k,n) = \mathbf{h}_{\mathrm{dir}}(k,n)\, G_i^*(k,n), \quad (7)$$

onde

$$\mathbf{h}_{\mathrm{dir}}(k,n) = \mathbf{\Phi}_{\mathrm{u}}^{-1}(k,n)\mathbf{a}(k,\varphi)\left[\mathbf{a}^{\mathrm{H}}(k,\varphi)\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{a}(k,\varphi)\right]^{-1}. \quad (8)$$

[170] O cálculo do filtro requer que o vetor de propagação de matriz **a**(*k*, *φ*),

que pode ser determinado após a DOA $\varphi(k, n)$ do som direto foi estimado [8]. Como explicado acima, o vetor de propagação de matriz e, desse modo, o filtro depende da DOA. A DOA pode ser estimada tal como se explica abaixo.

[171] O filtro espacial informado proposto em [8], por exemplo, a extração de som direto usando (4) e (7), não pode ser usado diretamente na modalidade na Fig. 2. De fato, o cálculo requer os sinais de microfone $\mathbf{x}(k, n)$ bem como o ganho de som direto $G_i(k, n)$. Como pode ser visto na Fig. 2, os sinais de microfone $\mathbf{x}(k, n)$ só estão disponíveis no lado de extremidade próxima, enquanto o ganho de som direto $G_i(k, n)$ só está disponível no lado de extremidade distante.

[172] A fim de usar o filtro espacial informado nas modalidades da invenção, uma modificação é fornecida, em que nós substituímos (7) em (4), levando a

$$\widehat{Y}_{\mathrm{dir},i}(k,n) = G_i(k,n)\widehat{X}_{\mathrm{dir}}(k,n), \quad (9)$$

onde

$$\widehat{X}_{\mathrm{dir}}(k,n) = \mathbf{h}_{\mathrm{dir}}^{\mathrm{H}}(k,n)\mathbf{x}(k,n). \quad (10)$$

[173] Este filtro modificado $\mathbf{h}_{\mathrm{dir}}(k, n)$ é independente das ponderações $G_i(k, n)$. Deste modo, o filtro pode ser aplicado no lado de extremidade próxima para obter o som direto $\hat{X}_{dir}(k,n)$, que podem então ser transmitidos para o lado de extremidade distante em conjunto com as DOAs estimadas (e distância), como informação lateral para fornecer um controle total sobre a reprodução do som direto. O som direto $\hat{X}_{dir}(k,n)$ pode ser determinado com respeito a um microfone de referência em uma posição $\mathbf{d}_1$. Portanto, pode-se também relacionar os componentes de som direto como $\hat{X}_{dir}(k,n,\mathbf{d}_1)$, e, portanto:

$$\widehat{X}_{\mathrm{dir}}(k,n,\mathbf{d}_1) = \mathbf{h}_{\mathrm{dir}}^{\mathrm{H}}(k,n)\mathbf{x}(k,n). \quad (10a)$$

[174] Desse modo, de acordo com uma modalidade, o módulo de decomposição 101 pode, por exemplo, ser configurado para gerar o sinal de componente direto através da aplicação de um filtro sobre os dois ou mais sinais de entrada de áudio de acordo com

$$\widehat{X}_{\mathrm{dir}}(k,n) = \mathbf{h}_{\mathrm{dir}}^{\mathrm{H}}(k,n)\mathbf{x}(k,n),$$

[175] em que k indica a frequência, e em que n indica o tempo, em que $\hat{X}_{dir}(k,n)$ indica o sinal de componente direto, em que $\mathbf{x}(k, n)$ indica os dois ou

mais sinais de entrada de áudio, em que $\mathbf{h}_{dir}(k, n)$ indica o filtro, com

$$\mathbf{h}_{\mathrm{dir}}(k,n) = \mathbf{\Phi}_{\mathrm{u}}^{-1}(k,n)\mathbf{a}(k,\varphi)\left[\mathbf{a}^{\mathrm{H}}(k,\varphi)\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{a}(k,\varphi)\right]^{-1}.$$

[176] em que $\mathbf{\Phi}_u(k, n)$ indica uma matriz de densidade espectral de potência do ruído e som difuso de dois ou mais sinais de entrada de áudio, em que $\mathbf{a}(k, \varphi)$ indica um vetor de propagação de matriz, e em que φ indica o ângulo de azimute da direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio.

[177] A Fig. 3 ilustra o módulo de estimativa de parâmetro 102 e um módulo de decomposição 101 implementando a decomposição direta/difusa de acordo com uma modalidade.

[178] A modalidade ilustrada pela Fig. 3 realiza a extração de som direto através do módulo de extração de som direto 203 e difunde a extração de som através do módulo de extração de som 204.

[179] A extração de som direto é realizada no módulo de extração de som direto 203, através da aplicação das ponderações de filtro para os sinais de microfone como dado em (10). As ponderações de filtro direto são calculadas na unidade de computação de ponderações diretas 301, que pode ser realizada por exemplo, com (8). Os ganhos $G_i(k, n)$ de, por exemplo, a equação (9), são, em seguida, aplicados no lado de extremidade distante, como mostrado na Fig. 2.

[180] A seguir, a extração de som difuso é descrita. A extração de som difuso pode, por exemplo, ser implementada através do módulo de extração de som difuso 204 da Fig. 3. As ponderações de filtro difusas são calculadas na unidade de computação de ponderações difusas 302 da Fig. 3, por exemplo, como descrito a seguir.

[181] Nas modalidades, o som difuso pode, por exemplo, ser extraído usando o filtro espacial que foi recentemente proposto em [9]. O som difuso $X_{diff}(k, n)$ em (2a) e a Fig. 2 podem, por exemplo, ser estimados mediante a aplicação de um segundo filtro espacial para os sinais de microfone, por exemplo,

$$\widehat{X}_{\mathrm{diff}}(k,n) = \mathbf{h}_{\mathrm{diff}}^{\mathrm{H}}(k,n)\mathbf{x}(k,n). \quad (11)$$

[182] Para encontrar o filtro ideal para o som difuso $\mathbf{h}_{diff}(k, n)$, considera-se o filtro recentemente proposto em [9], que pode extrair o som difuso com uma

resposta arbitrária desejada, minimizando o ruído na saída do filtro. Para o ruído espacialmente branco, o filtro é dado por

$$\mathbf{h}_{\text{diff}}(k,n) = \arg\min_{\mathbf{h}} \mathbf{h}^{\text{H}}\mathbf{h} \quad (12)$$

[183] sujeito a $\mathbf{h}^{\text{H}}\mathbf{a}(k, \varphi) = 0$ e $\mathbf{h}^{\text{H}}\boldsymbol{\gamma}_1(k) = 1$. A primeira restrição linear assegura que o som direto é suprimido, enquanto que a segunda restrição assegura que, em média, o som difuso é capturado com o ganho Q desejado, ver documento [9]. Note-se que $\boldsymbol{\gamma}_1(k)$ é o vetor de coerência de som difuso definido em [9]. A solução (12) é dada por

$$\mathbf{h}_{\text{diff}}(k,n) = \boldsymbol{\Lambda}\boldsymbol{\gamma}_{\text{diff}}(k)\left[\boldsymbol{\gamma}_{\text{diff}}^{\text{H}}(k)\boldsymbol{\Lambda}\boldsymbol{\gamma}_{\text{diff}}(k)\right]^{-1}, \quad (13)$$

onde

$$\boldsymbol{\Lambda}(k,\varphi) = \mathbf{I} - \mathbf{a}(k,\varphi)\left[\mathbf{a}^{\text{H}}(k,\varphi)\mathbf{a}(k,\varphi)\right]^{-1}\mathbf{a}^{\text{H}}(k,\varphi) \quad (14)$$

[184] sendo $\mathbf{I}$ a matriz de identidade de tamanho $M \times M$. O filtro $\mathbf{h}_{\text{diff}}(k, n)$ não depende das ponderações $G_i(k, n)$ e Q, e, desse modo, pode ser calculado e aplicado no lado de extremidade próxima para obter $\hat{X}_{diff}(k,n)$. Ao fazê-lo, só é necessário transmitir um único sinal de áudio para o lado de extremidade distante, ou seja, sendo ainda capaz de controlar totalmente a reprodução de som espacial do som difuso.

[185] A Fig. 3 ilustra além disso, a extração de som difuso de acordo com uma modalidade. A extração de som difuso é realizada no módulo de extração de som difuso 204, aplicando as ponderações de filtro para os sinais de microfone como dado na fórmula (11). As ponderações de filtro são calculadas na unidade de computação de ponderações difusas 302, que pode ser realizada, por exemplo, através do emprego da fórmula (13).

[186] A seguir, a estimativa de parâmetros é descrita. A estimativa de parâmetro pode, por exemplo, ser conduzida pelo módulo de estimativa de parâmetro 102, em que as informações paramétricas sobre a cena de som gravada podem, por exemplo, ser estimadas. Estas informações paramétricas são empregadas para calcular dois filtros espaciais no módulo de decomposição 101 e para a seleção de ganho na reprodução de áudio espacial consistente no modificador de sinais 103.

[187] Primeiramente, a determinação/estimativa de informações de DOA é descrita.

[188] A seguir são descritas modalidades em que o módulo de estimativa de parâmetro (102) compreende um estimador de DOA para o som direto, por exemplo, para uma onda plana que se origina a partir da posição da fonte de som e chega na matriz de microfones. Sem perda de generalidade, presume-se que existe uma única onda plana para cada tempo e frequência. Outras modalidades consideram casos em que várias ondas planas existem, e a extensão dos conceitos de ondas planas individuais descritos aqui para múltiplas ondas planas é simples. Como consequência, a presente invenção também abrange modalidades com múltiplas ondas planas.

[189] As DOAs de banda estreita podem ser estimadas a partir dos sinais de microfone usando um dos estimadores de DOA banda estreita do estado da técnica, como o ESPRIT [10] ou root MUSIC [11]. Em vez do ângulo de azimute $\varphi(k, n)$, as informações de DOA também podem ser fornecidas sob a forma de frequência espacial $\mu[k \mid \varphi(k, n)]$, de alteração de fase, ou de vetor de propagação $\mathbf{a}[k \mid \varphi(k, n)]$ para uma ou mais ondas que chegam na matriz de microfones. Deve-se notar que as informações de DOA também podem ser fornecidas externamente. Por exemplo, a DOA da onda plana pode ser determinada por uma câmara de vídeo juntamente com um algoritmo de reconhecimento de face assumindo que locutores humanos formam a cena acústica.

[190] Finalmente, deve-se notar que as informações de DOA também podem ser estimadas em 3D (em três dimensões). Nesse caso, tanto o ângulo de azimute $\varphi(k, n)$ quanto o de elevação $\vartheta(k, n)$ são estimados no módulo de estimativa de parâmetro 102 e a ACD da onda plana é em tal caso, fornecida, por exemplo, como (φ, ϑ).

[191] Desse modo, quando é feita referência abaixo ao ângulo de azimute de DOA, entende-se que todas as explicações são também aplicáveis para o ângulo de elevação de DOA, para um ângulo ou derivada do ângulo de azimute de DOA, para um ângulo ou derivado do ângulo de elevação de DOA ou para um ângulo derivado do ângulo de azimute e do ângulo de elevação de DOA. De modo mais geral, todas as explicações dadas abaixo são igualmente aplicáveis a qualquer ângulo de acordo com a DOA.

[192] Agora, a determinação/estimativa de informações de distância é descrita.

[193] Algumas modalidades relacionam zoom acústico superior com base nas DOAs e nas distâncias. Em tais modalidades, o módulo de estimativa de parâmetro 102 pode, por exemplo, compreender dois sub-módulos, por exemplo, o sub-módulo estimador de DOA descrito acima e um sub-módulo de estimativa da distância que estima a distância a partir da posição de gravação para a fonte de som $r(k, n)$. Em tais modalidades, pode-se, por exemplo, assumir que cada onda plana que chega na matriz de microfones de gravação se origina a partir da fonte de som e propaga-se ao longo de uma linha reta para a matriz (que também é conhecida como o percurso de propagação direta).

[194] Existem várias abordagens do estado da técnica para a estimativa de distância usando sinais de microfone. Por exemplo, a distância para a fonte pode ser encontrada por cálculo das razões de energia entre os sinais de microfones, como descrito em [12]. Em alternativa, a distância para a fonte $r(k, n)$ em compartimentos acústicos (por exemplo, quartos) pode ser calculada com base na razão de sinal-para-difuso (SDR) estimada [13]. As estimativas de SDR podem então ser combinadas com o tempo de reverberação de um quarto (conhecido ou estimado usando métodos do estado da técnica) para o cálculo da distância. Para SDR elevada, a energia de som direto é elevada em comparação com o som difuso, que indica que a distância para a fonte é pequena. Quando o valor de SDR é baixo, a potência de som direto é fraca em comparação com a reverberação do quarto, o que indica uma grande distância para a fonte.

[195] Em outras modalidades, em vez de calcular/estimar a distância através do emprego de um módulo de computação de distância no módulo de estimativa de parâmetro 102, as informações de distância externa podem, por exemplo, ser recebidas, por exemplo, a partir do sistema visual. Por exemplo, as técnicas do estado da técnica usadas nesta visão podem, por exemplo, ser usadas de modo que possam fornecer as informações de distância, por exemplo, Tempo de voo (Time of Flight, ToF), visão estereoscópica, e a luz estruturada. Por exemplo, nas câmaras de ToF, a distância para a fonte pode ser calculada a partir do tempo de voo medido de um sinal de luz emitido por uma câmara e

percorrer para a fonte e de volta para o sensor da câmara. A visão estérea do computador, por exemplo, utiliza dois pontos de vista a partir dos quais a imagem visual é capturada para calcular a distância para a fonte.

[196] Ou, por exemplo, as câmaras de luz estruturada podem ser usadas, em que um padrão conhecido de pixels é projetado sobre uma cena visual. A análise de deformações após a projeção permite que o sistema visual estime a distância para a fonte. Deve-se notar que as informações de distância $r(k, n)$ para cada binário de tempo-frequência é necessária para a reprodução de cena de áudio consistente. Se as informações de distância são fornecidas externamente por um sistema visual, a distância para a fonte $r(k, n)$ que corresponde a DOA $\varphi(k, n)$, pode, por exemplo, ser selecionada como o valor de distância a partir do sistema visual que corresponde ao da direção particular $\varphi(k, n)$.

[197] A seguir, a reprodução de cena acústica consistente é considerada. Primeiramente, a reprodução de cena acústica baseada em DOAs é considerado.

[198] A cena reprodução de cena acústica pode ser conduzida de tal forma que seja consistente com a cena acústica gravada. Ou, a reprodução de cena acústica pode ser conduzida de tal forma que é consistente com uma imagem visual. As informações visuais correspondentes podem ser fornecidas para obter consistência com uma imagem visual.

[199] A consistência pode, por exemplo, ser obtida por ajustar as ponderações $G_i(k, n)$ e Q em (2a). De acordo com as modalidades, o modificador de sinais 103, que pode, por exemplo, existir, no lado de extremidade próxima, ou, como mostrado na Fig. 2, no lado de extremidade distante, pode, por exemplo, receber os sons direto $\hat{X}_{dir}(k,n)$ e difuso $\hat{X}_{diff}(k,n)$ como entrada, em conjunto com as estimativas de DOA $\varphi(k, n)$ como informações laterais. Com base nessas informações recebidas, os sinais de saída $Y_i(k, n)$ para um sistema de reprodução disponível podem, por exemplo, ser gerados, por exemplo, de acordo com a fórmula (2a).

[200] Em algumas modalidades, os parâmetros $G_i(k, n)$ e Q são selecionados nas unidades de seleção de ganho 201 e 202, respectivamente, a partir de duas funções de ganho $g_i(\varphi(k, n))$ e $q(k, n)$ fornecidas pelo módulo de computação de função de ganho 104.

[201] De acordo com uma modalidade, $G_i(k, n)$ pode, por exemplo, ser selecionada apenas com base nas informações de DOA e Q pode, por exemplo, ter um valor constante. Em outras modalidades, no entanto, outra ponderação $G_i(k, n)$ pode, por exemplo, ser determinada com base em informações adicionais, e a ponderação Q pode, por exemplo, ser variavelmente determinada.

[202] Primeiramente, são consideradas implementações que constatam a consistência com a cena acústica gravada. A seguir, são consideradas modalidades que constatam a consistência com as informações da imagem/com uma imagem visual sendo considerada.

[203] A seguir, um cálculo das ponderações $G_i(k, n)$ e Q é descrito para reproduzir uma cena acústica que é consistente com a cena acústica gravada, por exemplo, de tal modo que o ouvinte posicionado em um melhor local do sistema de reprodução, percebe as fontes de som como chegando a partir de DOAs das fontes de som na cena de som gravado, tendo a mesma potência que a cena gravada, e reproduzir a mesma percepção do som difuso circundante.

[204] Para uma configuração de alto-falantes conhecida, a reprodução da fonte de som a partir da direção $\varphi(k, n)$ pode, por exemplo, ser conseguida selecionando o ganho de som direto $G_i(k, n)$ na unidade de seleção de ganho 201 ("Direct Gain Selection - Seleção de Ganho Direto") a partir de uma tabela de pesquisa fixa fornecida pelo módulo de computação de função de ganho 104 para a DOA $\varphi(k, n)$ estimada, que pode ser escrito como

$$G_i(k,n) = g_i(\varphi(k,n)), \quad (15)$$

[205] onde $g_i(\varphi) = p_i(\varphi)$ é uma função que retorna o ganho de panning em todas as DOAs para o i-ésimo alto-falante. A função de ganho de panning $p_i(\varphi)$ depende da configuração do alto-falante e do esquema de panning.

[206] Um exemplo da função de ganho de panning $p_i(\varphi)$ tal como definido pelo panning de amplitude de base de vetor (vector base amplitude panning, VBAP) [14] para o alto-falante esquerdo e direito na reprodução de som estéreo é mostrado na Fig. 5 (a).

[207] Na Fig. 5 (a), um exemplo de uma função de ganho de panning VBAP $p_{b,i}$ para uma configuração de estéreo é ilustrado, e na Fig. 5 (b) os ganhos de panning para reprodução consistente são ilustrados.

[208] Por exemplo, se o som direto chega a partir de $\varphi(k, n) = 30^\circ$, o ganho do alto-falante à direita é $G_r(k, n) = g_r(30^\circ) = p_r(30^\circ) = 1$ e o ganho do alto-falante à esquerda é $G_l(k, n) = g_l(30^\circ) = p_l(30^\circ) = 0$. Para o som direto que chega de $\varphi(k, n) = 0^\circ$, os ganhos finais de alto-falantes estéreos são $G_r(k, n) = G_l(k, n) = \sqrt{0{,}5}$

[209] Em uma modalidade, a função de ganho de panning, por exemplo, $p_i(\varphi)$, pode, por exemplo, ser uma função de transferência relativa à cabeça (HRTF) em caso de reprodução de som binaural.

[210] Por exemplo, se HRTF $g_i(\varphi) = p_i(\varphi)$ retorna os valores complexos, então, o ganho de som direto $G_i(k, n)$ selecionado na unidade de seleção de ganho 201 pode, por exemplo, ser de valor complexo.

[211] Se três ou mais sinais de saída de áudio são gerados, os conceitos de panning do estado da técnica correspondente podem, por exemplo, ser usados para realização de panning de um sinal de entrada para os três ou mais sinais de saída de áudio. Por exemplo, VBAP para três ou mais sinais de saída de áudio poderá ser usado.

[212] Na reprodução de cena acústica consistente, a potência do som difuso deve permanecer a mesma que na cena gravada. Portanto, para o sistema de alto-falantes com, por exemplo, alto-falantes igualmente espaçados, o ganho de som difuso tem um valor constante:

$$Q = q_i = \frac{1}{\sqrt{I}}, \quad (16)$$

[213] em que I é o número de canais de alto-falante de saída. Isto significa que o módulo de computação de função de ganho 104 fornece um valor de saída único para o i-ésimo alto-falante (ou canal de fone de ouvido) dependendo do número de alto-falantes disponíveis para reprodução, e este valor é usado como o ganho Q difuso em todas as frequências. O som difuso final $Y_{diff,i}(k, n)$ para o i-ésimo canal de alto-falante é obtido por descorrelação $Y_{diff}(k, n)$, obtida em (2b).

[214] Desse modo, a reprodução de cena acústica que é consistente com a cena acústica gravada pode ser conseguida, por exemplo, através da determinação de ganhos para cada um dos sinais de saída de áudio dependendo, por exemplo, de uma direção de chegada, através da aplicação de uma pluralidade de ganhos determinados $G_i(k, n)$no sinal de som direto $\hat{X}_{dir}(k,n)$ para determinar uma pluralidade de componentes de sinal de saída direto

$\hat{Y}_{dir,i}(k,n)$, através da aplicação de ganho *Q* determinado no sinal de som difuso $\hat{X}_{diff}(k,n)$ para obter um componente de sinal de saída difuso $\hat{Y}_{diff}(k,n)$ e através da combinação de cada um da pluralidade de componentes de sinal de saída direto $\hat{Y}_{dir,i}(k,n)$ com o componente de sinal de saída difuso $\hat{Y}_{diff}(k,n)$para a obtenção de um ou mais sinais de saída de áudio $Y_i(k,n)$.

[215] Agora, a geração de sinal de saída de áudio de acordo com as modalidades que obtém consistência com a cena visual é descrita. Em particular, o cálculo das ponderações $G_i(k, n)$ e *Q* de acordo com modalidades que são usadas para a reprodução de uma cena acústica que é consistente com a cena visual é descrito. É destinado a recriar uma imagem acústica em que o som direto a partir de uma fonte é reproduzido a partir da direção em que a fonte é visível no vídeo/imagem.

[216] Uma geometria, como representado na Fig. 4 pode ser considerada, em que ***I*** corresponde à direção de busca da câmara visual. Sem perda de generalidade, ***I*** pode definir o eixo Y do sistema de coordenadas.

[217] O azimute de DOA do som direto no sistema de coordenadas representado (*x, y*) é dada por $\varphi(k, n)$ e a localização da fonte no eixo-x é dada por $x_g(k, n)$. Aqui, assume-se que todas as fontes de som estão localizadas na mesma distância g ao eixo-x, por exemplo, as posições de origem estão localizadas na linha tracejado à esquerda, que é chamado em óptica de um plano focal. Deve-se notar que esta hipótese é feita apenas para assegurar que as imagens visuais e acústicas estão alinhadas e o real valor da distância g não é necessária para o processamento apresentado.

[218] No lado da reprodução (lado de extremidade distante), o visor está localizado em *b* e a posição da fonte no visor é dada por $x_b(k, n)$. Além disso, x_d é o tamanho do visor (ou, em algumas modalidades, por exemplo, x_d indica a metade do tamanho do visor), φ_d é o ângulo visual máximo correspondente, S é o melhor local do sistema de reprodução de som, e $\varphi_b(k, n)$ é o ângulo a partir do qual o som direto deverá ser reproduzido de modo que as imagens visuais e acústicas são alinhadas. $\varphi_b(k, n)$ depende de $x_b(k, n)$ e da distância entre o melhor local *S* e o visor localizado em *b*. Além disso, $x_b(k, n)$ depende de vários parâmetros, tais como a distância *g* da fonte a partir da câmara, o tamanho do

sensor de imagem, e o tamanho do visor x_d. Infelizmente, pelo menos alguns destes parâmetros são, muitas vezes, desconhecidos, na prática, de tal modo que $x_b(k, n)$ e $\varphi_b(k, n)$ não podem ser determinados para uma dada DOA $\varphi_g(k, n)$. No entanto, assume-se que o sistema óptico é linear, de acordo com a fórmula (17):

$$\tan \varphi_b(k,n) = c \tan \varphi(k,n), \quad (17)$$

[219] onde *c* é uma constante desconhecida que compensa os parâmetros desconhecidos acima mencionados. Deve-se notar que *c* é constante apenas se todas as posições da fonte têm a mesma distância *g* ao eixo x.

[220] A seguir, *c* é assumido como sendo um parâmetro de calibração, que deve ser ajustado durante a fase de calibração até que as imagens visuais e acústicas sejam consistentes. Para realizar a calibração, as fontes de som devem ser posicionadas em um plano focal e o valor de *c* é encontrado de tal modo que as imagens visuais e acústicas são alinhadas. Uma vez calibrado, o valor de *c* permanece inalterado e o ângulo a partir do qual o som direto deve ser reproduzido é dado por

$$\varphi_b(k,n) = \tan^{-1}[c \tan \varphi(k,n)]. \quad (18)$$

[221] Para assegura que ambas as cenas acústica e visual são consistentes, a função de panning original $p_i(\varphi)$ é modificada para uma função de panning consistente (modificada), $p_{b,i}(\varphi)$. O ganho som direto $G_i(k, n)$ é agora selecionado de acordo com

$$G_i(k,n) = g_i(\varphi(k,n)), \quad (19)$$

$$g_i(\varphi) = p_{b,i}(\varphi), \quad (20)$$

[222] onde $p_{b,i}(\varphi)$ é a função de panning consistente que retorna os ganhos de panning para o *i*-ésimo alto-falante em todas as possíveis fontes de DOAs. Para um valor fixo de *c*, tal função de panning consistente é calculada no módulo de computação de função de ganho 104 a partir da tabela de ganho de panning original (por exemplo, VBAP) como

$$p_{b,i}(\varphi) = p_i(\tan^{-1}[c \tan \varphi]). \quad (21)$$

[223] Desse modo, nas modalidades, o processador de sinais 105 pode, por exemplo, ser configurado para determinar, para cada sinal de saída de áudio de

um ou mais sinais de saída de áudio, de tal modo que o ganho direto $G_i(k, n)$ é definido de acordo com a

$$G_i(k, n) = p_i(\tan^{-1}[c \tan(\varphi(k, n))]) .$$

[224] em que i indica um índice do dito sinal de saída de áudio, em que k indica a frequência, e em que n indica o tempo, em que $G_i(k, n)$ indica o ganho direto, em que $\varphi(k, n)$ indica um ângulo dependendo da direção de chegada (por exemplo, o ângulo de azimute da direção de chegada), em que c indica um valor constante, e em que p_i indica uma função de panning.

[225] Nas modalidades, o ganho de som direto $G_i(k, n)$ é selecionado na unidade de seleção de ganho 201 com base na DOA $\varphi(k, n)$ estimada a partir de uma tabela de pesquisa fixa fornecida pelo módulo de computação de função de ganho 104, qual é calculado uma vez (após a fase de calibração) usando (19).

[226] Desse modo, de acordo com uma modalidade, o processador de sinais 105 pode, por exemplo, ser configurado para obter, para cada sinal de saída de áudio um ou mais sinais de saída de áudio, o ganho direto para o dito sinal de saída áudio a partir de uma tabela de pesquisa dependendo da direção de chegada.

[227] Em uma modalidade, o processador de sinais 105 calcula uma tabela de pesquisa para a função de ganho direto $g_i(k, n)$. Por exemplo, para cada grau completo possível, por exemplo, 1°, 2°, 3°,..., para o valor do azimute φ de DOA, o ganho direto $G_i(k, n)$ pode ser calculado e armazenado com antecedência. Então, quando um valor de azimute φ atual da direção de chegada é recebido, o processador de sinais 105 lê o ganho direto $G_i(k, n)$ para o valor do azimute atual φ da tabela de pesquisa. (O valor de azimute φ atual, pode, por exemplo, ser o valor de argumento da tabela de pesquisa, e o ganho direto $G_i(k, n)$ pode, por exemplo, ser o valor de retorno da tabela de pesquisa). Em vez do azimute φ de DOA, em outras modalidades, a tabela de pesquisa pode ser calculada para qualquer ângulo dependendo da direção de chegada. Isto tem uma vantagem, de que o valor de ganho nem sempre tem que ser calculado para cada ponto no tempo, ou para cada binário de tempo-frequência, mas em vez disso, a tabela de pesquisa é calculada uma vez e, em seguida, para um ângulo recebido φ, o ganho direto $G_i(k, n)$ é lido a partir da tabela de pesquisa.

[228] Desse modo, de acordo com uma modalidade, o processador de

sinais 105 pode, por exemplo, ser configurado para calcular uma tabela de pesquisa, em que a tabela de pesquisa compreende uma pluralidade de entradas, em que cada uma das entradas compreende um valor de argumento da tabela de pesquisa e um valor de retorno da tabela de pesquisa sendo atribuídos ao dito valor de argumento. O processador de sinais 105 pode, por exemplo, ser configurado para obter um dos valores de retorno da tabela de pesquisa a partir da tabela de pesquisa selecionando um dos valores de argumento da tabela de pesquisa a partir da tabela de pesquisa, dependendo da direção de chegada. Além disso, o processador de sinais 105 pode, por exemplo, ser configurado para determinar o valor de ganho, pelo menos, um de um ou mais sinais de saída de áudio, dependendo do dito um dos valores de retorno da tabela de pesquisa obtidos a partir da tabela de pesquisa.

[229] O processador de sinais 105 pode, por exemplo, ser configurado para obter mais um dos valores de retorno da tabela de pesquisa (o mesmo) a partir da tabela de pesquisa, selecionando um outro dos valores de argumento da tabela de pesquisa, dependendo de outra direção de chegada para determinar um outro valor de ganho. Por exemplo, o processador de sinais pode, por exemplo, receber informações de direção adicionais, por exemplo, em um ponto no tempo mais tarde, o que depende da dita direção de chegada adicional.

[230] Um exemplo de funções de ganho de panning consistentes e o panning de VBAP são mostrados na Fig. 5(a) e 5(b).

[231] Deve-se notar que, em vez de recalcular as tabelas de ganho de panning, pode-se calcular, em alternativa, a DOA $\varphi_b(k, n)$ para o visor e aplicá-la na função do panning original como $\varphi_i(\varphi_b(k, n))$. Isto é verdade, uma vez que a seguinte relação ocorre:

$$p_{b,i}(\varphi(k,n)) = p_i(\varphi_b(k,n)). \quad (22)$$

[232] No entanto, isto requer que o módulo de computação de função de ganho 104 receba também as DOAs $\varphi(k, n)$ estimadas como entrada e o novo cálculo de DOA, por exemplo, conduzido de acordo com a fórmula (18), que, em seguida, pode ser realizada para cada índice de tempo n.

[233] No que se refere à reprodução de som difuso, as imagens acústicas e visuais são consistentemente recriadas quando processadas da mesma

maneira como explicado para o caso sem os efeitos visuais, por exemplo, quando a potência do som difuso permanece a mesma que a potência difusa na cena gravada e os sinais de alto-falantes são versões não correlacionadas de $Y_{diff}(k, n)$. Para alto-falantes igualmente espaçados, o ganho de som difuso tem um valor constante, por exemplo, dado pela fórmula (16). Como resultado, o módulo de computação de função de ganho 104 fornece um valor de saída único para o *i*-ésimo alto-falante (ou canal de fone de ouvido) que é usado como o ganho difuso Q em todas as frequências. O som difuso final $Y_{diff,i}(k, n)$ para o *i*-ésimo canal de alto-falante é obtido por descorrelação de $Y_{diff}(k, n)$, por exemplo, como dado pela fórmula (2b).

[234] Agora, são consideradas as modalidades em que é fornecido um zoom acústico baseado em DOAs. Em tais modalidades, o processamento de um zoom acústico pode ser considerado o qual é consistente com o zoom visual. Este zoom audiovisual consistente é conseguido ajustando as ponderações $G_i(k, n)$ and Q, por exemplo, usadas na fórmula (2a) como descrito no modificador de sinais 103 da Fig. 2.

[235] Em uma modalidade, o ganho direto $G_i(k, n)$ pode, por exemplo, ser selecionado na unidade de seleção de ganho 201 a partir da função de ganho direto $g_i(k, n)$ calculada no módulo de computação de função de ganho 104 com base em DOAs estimada em módulo de estimativa de parâmetro 102. O ganho difuso Q é selecionado na unidade de seleção de ganho 202 a partir da função de ganho difuso $q(\beta)$ calculada em módulo de computação de função de ganho 104. Em outras modalidades, o ganho direto $G_i(k, n)$ e o ganho difuso Q são calculados pelo modificador de sinais 103 sem calcular primeiro, as respectivas funções de ganho e, em seguida, selecionar os ganhos.

[236] Deve-se notar que, em contraste com a modalidade acima descrita, a função de ganho difuso $q(\beta)$ é determinada com base no fator de zoom β. Nas modalidades, as informações sobre a distância não são usadas, e desse modo, em tais modalidades, elas não são calculadas no módulo de estimativa de parâmetro 102.

[237] Para derivar os parâmetros de zoom $G_i(k, n)$ e Q em (2a), a geometria da Fig. 4 é considerada. Os parâmetros indicados na figura são análogos aos descritos em relação à Fig. 4 na modalidade acima.

[238] De modo semelhante à modalidade acima descrita, presume-se que todas as fontes de som estão localizadas no plano focal, que é posicionado paralelo ao eixo x, a uma distância g. Deve-se notar que alguns sistemas de autofocagem são capazes de fornecer *g*, por exemplo, a distância ao plano focal. Isto permite supor que todas as fontes na imagem são nítidas. Na reprodução o lado (extremidade distante), a DOA $\varphi_b(k, n)$ e a posição $x_b(k, n)$ em um visor depende de muitos parâmetros, tais como a distância *g* da fonte a partir da câmara, o tamanho do sensor de imagem, o tamanho do visor x_d, e o fator de zoom da câmera (por exemplo, o ângulo da câmera de abertura) β. Assumindo-se que o sistema óptico é linear, de acordo com a fórmula (23):

$$\tan \varphi_b(k,n) = \beta\, c\, \tan \varphi(k,n), \quad (23)$$

[239] em que *c* é o parâmetro de calibração para compensar os parâmetros ópticos desconhecidos e $\beta \geq 1$ é o fator de zoom controlado pelo usuário. Deve-se notar que, em uma câmara visual, o aumento do zoom, por um fator β é equivalente a multiplicar $x_b(k, n)$ por β. Além disso, *c* é constante apenas se todas as posições da fonte têm a mesma distância *g* ao eixo x. Neste caso, *c* pode ser considerada como um parâmetro de calibração que é ajustado uma vez de tal modo que as imagens visuais e acústicas são alinhadas. O ganho de som direto $G_i(k, n)$ é selecionado a partir da função de ganho direto $g_i(\varphi)$ como

$$G_i(k,n) = g_i(\varphi(k,n)), \quad (24)$$

$$g_i(\varphi) = p_{b,i}(\varphi)\; w_b(\varphi), \quad (25)$$

[240] onde $p_{b,i}(\varphi)$ denota a função de ganho de panning e $w_b(\varphi)$ é a função de janela de ganho para um zoom áudio-visual consistente. A função de ganho de panning de um zoom áudio-visual consistente é calculada no módulo de computação de função de ganho 104 a partir da função de ganho de panning original (por exemplo, VBAP) $p_i(\varphi)$ como

$$p_{b,i}(\varphi) = p_i(\tan^{-1}[\beta\, c\, \tan \varphi]). \quad (26)$$

[241] Desse modo, o ganho de som direto $G_i(k, n)$, por exemplo, selecionado na unidade de seleção de ganho 201, é determinado com base em DOA $\varphi(k, n)$ estimada a partir de uma tabela de panning de pesquisa calculada no módulo de computação de função ganho 104, que é fixa, se β não se altera. Deve-se notar que, em algumas modalidades, $p_{b,i}(\varphi)$ precisa de ser calculado

novamente, por exemplo, através do emprego de fórmula (26) toda vez que o fator de zoom β for modificado.

[242] Exemplo funções de ganho de panning estéreo para $\beta = 1$ e $\beta = 3$ são mostrados na Fig. 6 (ver Fig. 6(a) e a Fig. 6(b)). Em particular, a Fig. 6(a) ilustra um exemplo de função de ganho de panning de $p_{b,i}$ para $\beta = 1$; a Fig. 6 (b) ilustra ganhos de panning depois do zoom com $\beta = 3$; e a Fig. 6 (c) ilustra ganhos de panning depois do zoom com $\beta = 3$ com um deslocamento angular.

[243] Como pode ser visto no exemplo, quando o som direto chega a partir $\varphi(k, n) = 10°$, o ganho de panning para o alto-falante à esquerda é aumentado para grandes valores de β, enquanto que a função de panning para o alto-falante à direita e $\beta = 3$ retorna um valor menor do que para $\beta = 1$. Tal panning move efetivamente a posição de fonte percebida mais para as direções exteriores quando o fator de zoom β é aumentado.

[244] De acordo com modalidades, o processador de sinais 105 pode, por exemplo, ser configurado para determinar dois ou mais sinais de saída de áudio. Para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio, uma função de ganho de panning é atribuída ao dito sinal de saída de áudio.

[245] A função de ganho de panning de cada um de dois ou mais sinais de saída de áudio compreende uma pluralidade de valores de argumento de função de panning, em que um valor de retorno de função de panning é atribuído a cada um dos ditos valores de argumento de função de panning, em que, quando a dita função de panning recebe um dos ditos valores de argumento de função de panning, a dita função de panning é configurada para retornar o valor de retorno de função de panning sendo atribuído a um dos ditos valores de argumento de função de panning. E:

[246] O processador de sinais 105 é configurado para determinar cada um de dois ou mais sinais de saída de áudio dependendo de um valor de argumento dependente de direção dos valores de argumento de função de panning da função de ganho de panning sendo atribuída ao dito sinal de saída de áudio, em que o dito valor de argumento dependente de direção depende na direção de chegada.

[247] De acordo com uma modalidade, a função de ganho de panning de cada um de dois ou mais sinais de saída de áudio tem um ou mais máximos

globais, sendo um dos valores de argumento de função de panning, em que para cada um de um ou mais máximos globais de cada função de ganho de panning, não existe nenhum outro valor de argumento de função de panning para o qual a dita função de ganho de panning retorna um valor de retorno de função de panning maior do que para o dito máximo global.

[248] Para cada par de um primeiro sinal de saída de áudio e um segundo sinal de saída de áudio de dois ou mais sinais de saída de áudio, pelo menos um de um ou mais máximos globais da função de ganho de panning do primeiro sinal de saída de áudio é diferente de qualquer um de um ou mais máximos globais da função de ganho de panning do segundo sinal de saída de áudio.

[249] Dito em suma, as funções de panning são implementadas de tal forma que (pelo menos uma de) os máximos globais das diferentes funções panning se diferem.

[250] Por exemplo, na Fig. 6 (a), os máximos locais de $p_{b,l}(\varphi)$ estão na faixa de -45° a -28° e os máximos locais de $p_{b,r}(\varphi)$ estão na faixa de +28° a +45° e, portanto, os máximos globais são diferentes.

[251] Por exemplo, na Fig. 6(b), os máximos locais de $p_{b,l}(\varphi)$ estão na faixa de -45° a -8° e os máximos locais de $p_{b,r}(\varphi)$ estão na faixa de + 8° e + 45° e, portanto, a máximos globais são também diferentes.

[252] Por exemplo, na Fig. 6(c), os máximos locais de $p_{b,l}(\varphi)$ estão na faixa de -45° a +2° e os máximos locais de $p_{b,r}(\varphi)$ estão na faixa de +18° a +45° e, portanto, a máximos globais também diferem.

[253] A função de ganho de panning pode, por exemplo, ser implementada como uma tabela de pesquisa.

[254] Em tal modalidade, o processador de sinais 105 pode, por exemplo, ser configurado para calcular uma tabela de pesquisa de panning para uma função de ganho de panning de pelo menos um dos sinais de saída de áudio.

[255] A tabela de pesquisa de panning de cada sinal de saída de áudio do dito, pelo menos, um dos sinais de saída de áudio pode, por exemplo, compreender uma pluralidade de entradas, em que cada uma das entradas compreende um valor de argumento de função de panning da função de ganho de panning do dito sinal de saída de áudio e o valor de retorno de função de panning da função de ganho de panning sendo atribuído ao dito valor de

argumento de função de panning, em que o processador de sinais 105 é configurado para obter um dos valores de retorno de função de panning a partir da dita tabela de pesquisa de panning selecionada, dependendo da direção de chegada, o valor de argumento dependente de direção a partir da tabela de pesquisa de panning, e em que o processador de sinais 105 é configurado para determinar o valor de ganho para o dito sinal de saída de áudio de acordo com o dito um dos valores de retorno de função de panning obtidos a partir da dita tabela de pesquisa de panning.

[256] A seguir, são descritas modalidades que empregam uma janela de som direto. De acordo com tais modalidades, uma janela de som direto para o zoom consistente $w_b(\varphi)$ é calculada de acordo com a:

$$w_b(\varphi) = w(\tan^{-1}[\beta\, c\, \tan\varphi]), \quad (27)$$

[257] onde $w_b(\varphi)$ é uma função de janela de ganho para um zoom acústico que atenua o som direto, se a fonte é mapeada para uma posição fora da imagem visual para o fator de zoom β.

[258] A função de janela $w(\varphi)$ pode, por exemplo, ser ajustada para $\beta = 1$, de tal modo que o som direto de fontes que estão fora da imagem visual é reduzido para um nível desejado, e pode ser calculado, por exemplo, empregando a fórmula (27), toda vez que os parâmetros de zoom se alteram. Deve-se notar que $w_b(\varphi)$ é o mesmo para todos os canais de alto-falantes. Exemplos funções de janela para $\beta = 1$ e $\beta = 3$ são mostrados na Fig. 7(a-b), em que para um valor de β aumentado da largura da janela é diminuída.

[259] Na Fig. 7 exemplos de funções de janela de ganho consistentes são ilustrados. Em particular, a Fig. 7(a) ilustra uma função de janela de ganho w_b sem zoom (fator de zoom $\beta = 1$), a Fig. 7 (b) ilustra uma função de janela de ganho depois do zoom (fator de zoom $\beta = 3$), a Fig. 7 (c) ilustra uma função de janela de ganho depois do zoom (fator de zoom $\beta = 3$) com um desvio angular. Por exemplo, o desvio angular pode realizar uma rotação da janela, para uma direção de busca.

[260] Por exemplo, na Fig. 7(a), 7(b) e 7(c), a função de janela de ganho retorna um ganho de 1, se a DOA φ está localizada dentro da janela, a função de janela de ganho retorna um ganho de 0,18, se φ está localizado fora da janela,

e a função de janela de ganho retorna um ganho entre 0,18 e 1, se φ está localizado na fronteira da janela.

[261] De acordo com modalidades, o processador de sinais 105 é configurado para gerar cada sinal de saída de áudio do um ou mais sinais de saída de áudio dependendo de uma função de janela de ganho. A função de janela de ganho é configurada para retornar um valor de retorno de função de janela ao receber um valor de argumento de função de janela.

[262] Se o valor de argumento de função de janela é maior do que um limite de janela inferior e menor do que um limite de janela superior, a função de janela de ganho é configurada para retornar um valor de retorno de função de janela sendo maior do que qualquer valor de retorno de função de janela retornado pela função ganho de janela, se o valor de argumento de função de janela é menor do que o limite inferior, ou maior do que o limite superior.

[263] Por exemplo, na fórmula (27)

$$w_b(\varphi) = w(\tan^{-1}[\beta\, c\, \tan\varphi]),$$

[264] o ângulo de azimute da direção de chegada φ é o valor de argumento de função de janela da função de janela de ganho. A função de janela de ganho depende das informações de zoom, aqui, fator de zoom β.

[265] Para explicar a definição da função de janela de ganho, pode ser feita referência à Fig. 7 (a).

[266] Se o ângulo de azimute de DOA φ é maior do que -20° (limite inferior) e menor do que +20° (limite superior), todos os valores retornados pela função de janela de ganho são maiores do que 0,6. Caso contrário, se o ângulo de azimute de DOA φ é menor do que -20° (limiar inferior) ou maior do que +20° (limite superior), todos os valores retornados pela função de janela de ganho são menores do que 0,6.

[267] Em uma modalidade, o processador de sinais 105 é configurado para receber informações de zoom. Além disso, o processador de sinais 105 é configurado para gerar cada sinal de saída de áudio de um ou mais sinais de saída de áudio dependendo da função de janela de ganho, em que a função de janela de ganho depende das informações de zoom.

[268] Isto pode ser visto para as funções de janela de ganho (modificadas) da Fig. 7(b) e da Fig. 7(c) se outros valores são considerados como limites

inferiores/superiores ou se outros valores são considerados como valores de retorno. Nas Figs. 7(a), 7(b) e 7(c), pode-se ver, que a função de janela de ganho depende das informações de zoom: fator de zoom β.

[269] A função de janela de ganho pode, por exemplo, ser implementada como uma tabela de pesquisa. Em tal modalidade, o processador de sinais 105 é configurado para calcular uma tabela de pesquisa de janela, em que a tabela de pesquisa de janela compreende uma pluralidade de entradas, em que cada uma das entradas compreende um valor de argumento de função de janela da função de janela de ganho e um valor de retorno de função de janela da função de janela de ganho sendo atribuída ao dito valor de argumento de função de janela. O processador de sinais 105 é configurado para obter um dos valores de retorno de função de janela da tabela de pesquisa de janela, selecionando um dos valores de argumento de função de janela de tabela de pesquisa de janela, dependendo da direção de chegada. Além disso, o processador de sinais 105 é configurado para determinar o valor de ganho para pelo menos um de um ou mais sinais de saída de áudio, dependendo do dito um dos valores de retorno de função de janela obtidos a partir da tabela de pesquisa de janela.

[270] Em adição ao conceito de zoom, as funções de janela e de panning podem ser deslocadas por um ângulo de deslocamento θ. Este ângulo poderia corresponder tanto para a rotação de uma direção de busca da câmera *I* quanto para movimento em uma imagem visual por analogia a um zoom digital nas câmeras. No primeiro caso, o ângulo de rotação da câmara é recalculado para o ângulo de um visor, por exemplo, de forma semelhante à fórmula (23). Neste último caso, θ pode ter um desvio direto das funções de janela e de panning (por exemplo, $w_b(\varphi)$ e $p_{b,i}(\varphi)$) para o zoom acústico consistente. Um exemplo ilustrativo de uma alteração de ambas as funções está representado nas Figs. 5(c) e 6(c).

[271] Deve-se notar que, em vez de recalcular as funções de ganho e de janela de panning, pode-se calcular a DOA $\varphi_b(k, n)$ para o visor, por exemplo, de acordo com a fórmula (23), e aplicá-la nas funções de panning e de janela originais como $p_i(\varphi)$ e $w(\varphi_b)$, respectivamente. O dito processamento é equivalente desde que as seguintes relações ocorram:

$$p_{b,i}(\varphi(k,n)) = p_i(\varphi_b(k,n)), \quad (28)$$
$$w_b(\varphi(k,n)) = w(\varphi_b(k,n)). \quad (29)$$

[272] No entanto, isto requer que o módulo de computação de função de ganho 104 receba as DOAs $\varphi(k, n)$ estimadas como entrada e o novo cálculo de DOA, por exemplo, de acordo com a fórmula (18), pode, por exemplo, ser realizado em cada período de tempo consecutivo, independentemente se β foi alterado ou não.

[273] Quanto ao som difuso, o cálculo da função de ganho difuso $q(\beta)$, por exemplo, no módulo de computação de função de ganho 104, requer apenas o conhecimento do número de alto-falantes I disponíveis para reprodução. Desse modo, ele pode ser configurado de forma independente a partir dos parâmetros de uma câmara visual ou do visor.

[274] Por exemplo, para alto-falantes igualmente espaçados, o ganho de som difuso de valor real $Q \in [0, 1/\sqrt{I}]$ na fórmula (2a) é selecionado na unidade de seleção de ganho 202 com base no parâmetro zoom β. O objetivo de usar o ganho difuso é atenuar o som difuso dependendo do fator de zoom, por exemplo, o zoom aumenta a DRR do sinal reproduzido. Isto é conseguido através da redução de Q para β maior. Na verdade, o aumento do zoom significa que o ângulo de abertura da câmara se torna menor, por exemplo, uma correspondência acústica natural seria um microfone mais diretivo que capture som menos difuso.

[275] Para imitar este efeito, uma modalidade pode, por exemplo, empregar a função de ganho mostrada na Fig. 8. A Fig. 8 ilustra um exemplo de uma função de ganho difuso $q(\beta)$.

[276] Em outras modalidades, a função de ganho é definida de forma diferente. O som difuso final $Y_{diff,i}(k, n)$ para o i-ésimo canal de alto-falante é conseguido por descorrelação de $Y_{diff}(k, n)$, por exemplo, de acordo com a fórmula (2b).

[277] A seguir, o zoom acústico com base em DOAs e nas distâncias é considerado.

[278] De acordo com algumas modalidades, o processador de sinais 105 pode, por exemplo, ser configurado para receber informações de distância, em

que o processador de sinais 105 pode, por exemplo, ser configurado para gerar cada sinal de saída de áudio dos uma ou mais sinais de áudio de saída, dependendo das informações de distância.

[279] Algumas modalidades empregam um processamento para o zoom acústico consistente que é determinado com base em DOA $\varphi(k, n)$ estimada e um valor de distância $r(k, n)$. Os conceitos destas modalidades podem também ser aplicados para alinhar a cena acústica gravada a um vídeo sem zoom, onde as fontes não estão localizadas na mesma distância como anteriormente assumido nas informações de distância $r(k, n)$ disponíveis nos permite criar um efeito de desfocagem acústico para as fontes de som que não aparecem acentuadas na imagem visual, por exemplo, para as fontes que não estão localizadas no plano focal da câmera.

[280] Para facilitar uma reprodução de som consistente, por exemplo, um zoom acústico, com desfocagem para fontes localizadas em diferentes distâncias, o ganho $G_i(k, n)$) e Q pode ser ajustado na fórmula (2a), como representado no modificador de sinais 103 da Fig. 2 com base em dois parâmetros estimados, a saber, $\varphi(k, n)$ e $r(k, n)$,, e dependendo do fator de zoom β. Se nenhum zoom for envolvido, β pode ser ajustado para $\beta = 1$.

[281] Os parâmetros $\varphi(k, n)$ e $r(k, n)$ podem, por exemplo, ser estimados no módulo de estimativa de parâmetro 102, como descrito acima. Nesta modalidade, o ganho direto $G_i(k, n)$ é determinado (por exemplo, sendo selecionado na unidade de seleção de ganho 201) com base na DOA e nas informações de distância a partir de uma ou mais funções de ganho direto gi,j(k, n) (que podem, por exemplo, ser calculada no módulo de computação de função de ganho 104). Do mesmo modo como descrito para as modalidades acima, o ganho difuso Q pode, por exemplo, ser selecionado na unidade de seleção de ganho 202 a partir da função de ganho difuso $q(\beta)$, por exemplo, calculado no módulo de computação de função de ganho 104 com base no fator de zoom β.

[282] Em outras modalidades, o ganho direto $G_i(k, n)$ e o ganho difuso Q são calculados pelo modificador de sinais 103 sem calcular primeiro as respectivas funções de ganho e, em seguida, selecionando os ganhos.

[283] Para explicar a reprodução de cena acústica e o zoom acústico para fontes de som a diferentes distâncias, é feita referência à Fig. 9. Os parâmetros

indicados na Fig. 9 são análogos aos descritos acima.

[284] Na Fig. 9, a fonte de som está localizada na posição P' a uma distância $R(k, n)$ em relação ao eixo x. A distância r, que pode, por exemplo, ser (k, n)-específica (específica de tempo-frequência: $r(k, n)$) indica a distância entre a posição da fonte e o plano focal (linha vertical à esquerda que passa através de g). Deve-se notar que alguns sistemas de autofocagem são capazes de fornecer g, por exemplo, a distância ao plano focal.

[285] A DOA do som direto a partir do ponto de vista da matriz de microfone é indicada por $\varphi'(k, n)$. Em contraste para outras modalidades, não se assume que todas as fontes estão localizadas à mesma distância g da lente da câmara. Desse modo, por exemplo, a posição P' pode ter uma distância arbitrária $R(k, n)$ para o eixo x.

[286] Se a fonte não está localizada no plano focal, a fonte aparecerá desfocada no vídeo. Além disso, as modalidades baseiam-se na verificação de que se a fonte está localizada em qualquer posição sobre a linha tracejada 910, ela aparecerá na mesma posição $x_b(k, n)$ no vídeo. No entanto, as modalidades baseiam-se na verificação de que a DOA $\varphi'(k, n)$ estimada do som direto irá se alterar caso a fonte se mova ao longo da linha tracejada 910. Em outras palavras, com base nas conclusões usadas para as modalidades, se a fonte se move paralela ao eixo y, a DOA $\varphi'(k, n)$ estimada pode variar enquanto x_b (e, portanto, a DOA $\varphi_b(k, n)$ a partir da qual o som deve ser reproduzido) permanece o mesmo. Como consequência, se a DOA $\varphi'(k, n)$ estimada é transmitida para o lado de extremidade distante e usada para a reprodução de som, como descrito nas modalidades anteriores, então, a imagem acústica e visual não estão mais alinhadas, se a fonte altera a sua distância $R(k, n)$.

[287] Para compensar este efeito e obter uma reprodução de som consistente, a estimativa de DOA, por exemplo, realizada no módulo de estimativa de parâmetro 102, estima a DOA do som direto como se a fonte estivesse localizada no plano focal na posição P. Esta posição representa a projeção de P' no plano focal. A DOA correspondente é indicada por $\varphi(k, n)$ na Fig. 9 e é usada no lado de extremidade distante para a reprodução de som consistente, da mesma forma que nas modalidades anteriores. A DOA $\varphi(k, n)$ (modificada) pode ser calculada a partir da DOA $\varphi'(k, n)$ estimada (original), com

base nas considerações geométricas, se *r* e *g* são conhecidos.

[288] Por exemplo, na Fig. 9, o processador de sinais 105 pode, por exemplo, calcular $\varphi(k, n)$ a partir de $\varphi'(k, n)$ *r* e *g* de acordo com:

$$\varphi = \arctan\left(\frac{\tan\varphi' \cdot (r+g)}{g}\right).$$

[289] Desse modo, de acordo com uma modalidade, o processador de sinais 105 pode, por exemplo, ser configurado para receber um ângulo de azimute original $\varphi'(k, n)$ da direção de chegada, sendo a direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio e é configurado para receber, adicionalmente, informações de distância, e pode, por exemplo, ser configurado para receber, adicionalmente, informações sobre a distância *r*. O processador de sinais 105 pode, por exemplo, ser configurado para calcular um ângulo de azimute modificado $\varphi(k, n)$, da direção de chegada, dependendo do ângulo de azimute da direção de chegada original $\varphi'(k, n)$ e dependendo das informações de distância *r* e *g*. O processador de sinais 105 pode, por exemplo, ser configurado para gerar cada sinal de saída de áudio de um ou mais dos sinais de saída de áudio, dependendo do ângulo de azimute da direção de chegada modificada $\varphi(k, n)$.

[290] As informações de distância necessárias podem ser estimadas como explicado acima (a distância *g* do plano focal pode ser obtida a partir do sistema de lentes ou das informações de autofocagem). Deve-se notar que, por exemplo, nesta modalidade, a distância $r(k, n)$ entre a fonte e o plano focal é transmitida para o lado de extremidade distante em conjunto com a DOA $\varphi(k, n)$ (mapeada).

[291] Além disso, por analogia com o zoom visual, as fontes que se encontram a uma grande distância *r* a partir do plano focal não aparecem nítidas na imagem. Este efeito é bem conhecido na óptica como a chamada profundidade de campo (depth-of-field, DOF), que define a faixa de distâncias de origem que aparecem aceitavelmente nítidas na imagem visual.

[292] Um exemplo de curva de DOF como função da distância *r* é representado na Fig. 10(a).

[293] A Fig. 10 ilustra exemplos de valores para a profundidade de campo (Fig. 10(a)), para uma frequência de corte de um filtro passa-baixo (Fig. 10(b)), e para o tempo de retardo em ms para o som direto repetido (Fig. 10 (c)).

[294] Na Fig. 10 (a), as fontes a uma pequena distância do plano focal ainda são nítidas, enquanto as fontes em distâncias maiores (ou mais perto ou mais longe da câmera) aparecem desfocadas. Desse modo, de acordo com uma modalidade, as fontes de som correspondentes são desfocadas tal que as imagens visuais e acústicas são consistentes.

[295] Para derivar os ganhos $G_i(k, n)$ e Q em (2a), que constatam a desfocagem acústica e a reprodução de som espacial consistente, o ângulo é considerado de modo que a fonte posicionada em $P(\varphi, r)$ vai aparecer em um visor. A fonte desfocada será exibida a

$$\tan \varphi_b(k,n) = \beta\, c\, \tan \varphi(k,n), \quad (30)$$

[296] em que c é o parâmetro de calibração, $\beta \geq 1$ é o fator de zoom controlado pelo usuário, $\varphi(k, n)$ é a DOA (mapeada), por exemplo, estimada no módulo de estimativa de parâmetro 102. Como mencionado antes, o ganho direto $G_i(k, n)$ em tais modalidades pode, por exemplo, ser calculado a partir de várias funções de ganho direto $g_{i,j}$. Em particular, duas funções de ganho $g_{i,1}(\varphi(k, n))$ e $g_{i,2}(r(k, n))$, podem, por exemplo, ser usadas em que a primeira função de ganho depende da DOA $\varphi(k, n)$, e em que a segunda função de ganho depende da distância $r(k, n)$. O ganho direto $G_i(k, n)$ pode ser calculado como:

$$G_i(k,n) = g_{i,1}(\varphi(k,n))\, g_{i,2}(r(k,n)) \quad (31)$$

$$g_{i,1}(\varphi) = p_{b,i}(\varphi)\, w_b(\varphi) \quad (32)$$

$$g_{i,2}(r) = b(r), \quad (33)$$

[297] onde $p_{b,i}(\varphi)$ indica a função de ganho de panning (para garantir que o som é reproduzido da direção à direita), em $w_b(\varphi)$ é a função de janela de ganho (para garantir que o som direto é atenuado se a fonte não é visível no vídeo), e em que $b(r)$ é a função de desfocagem (para desfocar as fontes acusticamente se elas não se encontram no plano focal).

[298] Deve-se notar que todas as funções de ganho podem ser definidas dependentes da frequência (que é aqui omitido por questões de brevidade). Deve ser ainda notado que nesta modalidade o ganho direto G_i é encontrado por seleção e multiplicação dos ganhos de duas funções de ganho diferentes, como mostrado na fórmula (32).

[299] Ambas as funções de ganho $p_{b,i}(\varphi)$ e $w_b(\varphi)$ são definidas de modo

análogo ao descrito acima. Por exemplo, elas podem ser calculadas, por exemplo, no módulo de computação de função de ganho 104, por exemplo, usando as fórmulas (26) e (27), e elas permanecem fixas, a menos que o fator de zoom β se altere. A descrição detalhada dessas duas funções foi fornecida acima. A função de desfocagem $b(r)$ retorna ganhos complexos que causam desfocagem, por exemplo, dispersão perceptível, uma fonte, e assim a função de ganho global g_i também tipicamente retornará um número complexo. Para simplicidade, a seguir, a desfocagem é indicada como uma função de uma distância em relação ao plano focal $b(r)$.

[300] O efeito de desfocagem pode ser obtido na forma de um selecionado ou uma combinação dos seguintes efeitos de desfocagem: filtragem passa-baixo, adição de som direto retardado, atenuação de som direto, suavização temporal e/ou dispersão de DOA. Desse modo, de acordo com uma modalidade, o processador de sinais 105 pode, por exemplo, ser configurado para gerar um ou mais sinais de saída de áudio através da condução de filtragem passa-baixo, ou através da adição de som direto retardado, ou através da condução de atenuação do som direto, ou através da condução de suavização temporal, ou através da condução de direção de dispersão de chegada.

[301] Filtragem de passa-baixo: Em visão, uma imagem visual não nítida pode ser obtida por filtragem de passa-baixo, a qual funde de forma eficaz os pixels vizinhos na imagem visual. Por analogia, um efeito de desfocagem acústico pode ser obtido por filtragem de passa-baixo do som direto com a frequência de corte selecionada com base na distância estimada da fonte para o plano focal r. Neste caso, a função de desfocagem $b(r, k)$ retorna os ganhos de filtro passa-baixo para frequência k e a distância r. Uma curva de exemplo para a frequência de corte de um filtro passa-baixo de primeira ordem para a frequência de amostragem de 16 kHz é mostrada na Fig. 10 (b). Para distâncias pequenas r, a frequência de corte é próxima da frequência de Nyquist, e, portanto, quase nenhuma filtragem passa-baixo é efetivamente realizada. Para valores de distâncias maiores, a frequência de corte é reduzida até que os níveis a 3 kHz, onde a imagem acústica é suficientemente desfocada.

[302] Adição de som direto retardado: A fim de desfocar a imagem acústica de uma fonte, podemos descorrelacionar o som direto, por exemplo, repetindo

uma atenuação do som direto depois de algum retardo τ (por exemplo, entre 1 e 30 ms). Esse tratamento pode, por exemplo, ser realizado de acordo com a função de ganho complexo de fórmula (34):

$$b(r,k) = 1 + \alpha(r)\, e^{-j\omega\tau(r)} \quad (34)$$

[303] onde α denota o ganho de atenuação do som repetido e τ é o retardo após o qual o som direto é repetido. Uma curva de retardo de exemplo (em ms) é mostrada na Fig. 10 (c). Para distâncias pequenas, o sinal retardado não é repetido e α é definido como zero. Para distâncias maiores, o retardo de tempo aumenta com a distância crescente, o que causa uma dispersão perceptível de uma fonte acústica.

[304] Atenuação do som direto: A fonte pode também ser vista como desfocada quando o som direto é atenuado por um fator constante. Neste caso $b(r)$ = const < 1. Como mencionado acima, a função de desfocagem $b(r)$ pode consistir em qualquer um dos efeitos de desfocagem mencionados ou como uma combinação destes efeitos. Além disso, o processamento alternativo que desfoca a fonte pode ser usado.

[305] Suavização temporal: A suavização do som direto através do tempo pode, por exemplo, ser usada para desfocar perceptivelmente a fonte acústica. Isto pode ser conseguido pela suavização do envelope do sinal direto extraído ao longo do tempo.

[306] Dispersão de DOA: Outro método para desfocar uma fonte acústica consiste em reproduzir o sinal da fonte a partir da faixa de direções em vez de apenas a direção estimada. Isto pode ser conseguido através da randomização do ângulo, por exemplo, tomando um ângulo aleatório de uma distribuição de Gauss centrado em torno de φ estimado. Aumentando-se a variância de uma tal distribuição, e, portanto, ampliando a faixa possível de DOA, aumenta-se a percepção de desfocagem.

[307] Analogicamente como descrito acima, o cálculo da função de ganho difuso $q(\beta)$ no módulo de computação de função de ganho 104, pode, em algumas modalidades, requerer apenas o conhecimento do número de alto-falantes I disponíveis para a reprodução. Desse modo, a função de ganho difuso $q(\beta)$ pode, em tais modalidades, ser ajustada como desejado para a aplicação.

Por exemplo, para alto-falantes igualmente espaçados, o ganho de som difuso $Q \in \left[0, 1/\sqrt{I}\right]$ de valor real na fórmula (2a) é selecionado na unidade de seleção de ganho 202 com base no parâmetro de zoom β. O objetivo de usar o ganho difuso é atenuar o som difuso dependendo do fator de zoom, por exemplo, o zoom aumenta DRR do sinal reproduzido. Isto é conseguido através da redução de Q para maior β. Na verdade, o zoom, significa que o ângulo de abertura da câmara se torna menor, por exemplo, uma correspondência acústica natural seria um microfone mais diretivo que capture som menos difuso. Para imitar este efeito, podemos utilizar, por exemplo, a função de ganho mostrada na Fig. 8. Claramente, a função de ganho poderia também ser definida de maneira diferente. Opcionalmente, o som difuso final $Y_{diff,i}(k, n)$ para o i-ésimo canal de alto-falante é obtido por descorrelação de $Y_{diff}(k, n)$ obtido na fórmula (2b).

[308] Agora, as modalidades são consideradas que constatam uma aplicação para os aparelhos auditivos e dispositivos de auxiliar de audição. Fig. 11 ilustra uma tal aplicação aparelho auditivo.

[309] Algumas modalidades estão relacionadas com aparelhos auditivos binaurais. Neste caso, presume-se que cada aparelho auditivo está equipado com pelo menos um microfone e que as informações possam ser trocadas entre os dois aparelhos auditivos. Devido a alguma perda auditiva, a pessoa com deficiência auditiva pode ter dificuldades com foco (por exemplo, concentração de sons vindos de um determinado ponto ou direção) em um som ou sons desejados. A fim de ajudar o cérebro da pessoa com deficiência auditiva a processar os sons que são reproduzidos por aparelhos auditivos, a imagem acústica é feita de acordo com o ponto ou direção de foco do usuário de aparelhos auditivos. É concebível que o ponto ou direção do foco seja predefinido, definido pelo usuário, ou definido por uma interface cérebro-máquina. Tais modalidades asseguram que os sons desejados (que são assumidos como chegando do ponto de foco ou direção de foco) e os sons indesejáveis aparecem espacialmente separados.

[310] Em tais modalidades, as direções dos sons diretos podem ser estimadas de diferentes modos. De acordo com uma modalidade, as indicações são determinadas com base nas diferenças de nível interaurais (inter-aural level differences, ILDs) e/ou diferenças de tempo interaural (inter-aural time

differences, ITDs) que são determinadas usando ambos os aparelhos auditivos (ver [15] e [16]).

[311] De acordo com outras modalidades, as direções dos sons diretos à esquerda e à direita são estimadas independentemente usando um aparelho auditivo, que está equipado com, pelo menos, dois microfones (ver [17]). As direções estimadas podem ser excitadas com base nos níveis de pressão sonora à esquerda e à direita do aparelho auditivo, ou na coerência espacial no aparelho auditivo à esquerda e à direita. Devido ao efeito de sombreamento de cabeça, diferentes estimadores podem ser usados para diferentes faixas de frequência (por exemplo, ILDs em altas frequências e ITDs em baixas frequências).

[312] Em algumas modalidades, os sinais de som direto e difuso podem, por exemplo, ser estimados usando as técnicas de filtragem espacial informadas acima mencionadas. Neste caso, os sons diretos e difusos conforme recebidos no auxílio de audição à esquerda e à direita podem ser avaliados separadamente (por exemplo, alterando o microfone de referência), ou os sinais de saída à esquerda e à direita podem ser gerados usando uma função de ganho para auxílio de audição à esquerda e à direita, respectivamente, de uma forma semelhante aos diferentes sinais de alto-falantes ou fones de ouvido são obtidos nas modalidades anteriores.

[313] A fim de separar espacialmente os sons desejados e indesejados, o zoom acústico explicado nas modalidades acima mencionadas pode ser aplicado. Neste caso, o ponto de foco ou a direção do foco determina o fator de zoom.

[314] Desse modo, de acordo com uma modalidade, um auxiliar auditivo ou um dispositivo de audição auxiliar pode ser fornecido, em que o auxiliar auditivo ou um dispositivo de audição auxiliar compreende um sistema tal como descrito acima, em que o processador de sinais 105 do sistema acima descrito determina o ganho direto para cada um de um ou mais sinais de saída de áudio, por exemplo, de acordo com uma direção de foco ou um ponto de foco.

[315] Em uma modalidade, o processador de sinais 105 do sistema acima descrito pode, por exemplo, ser configurado para receber informações de zoom. O processador de sinais 105 do sistema acima descrito pode, por exemplo, ser configurado para gerar cada sinal de saída de áudio de um ou mais sinais de

saída de áudio dependendo de uma função de janela de ganho, em que a função de janela de ganho depende das informações de zoom. Os mesmos conceitos tal como explicados com referência às Figs. 7(a), 7(b) e (c) são usados.

[316] Se um argumento da função de janela, dependendo da direção de foco ou do ponto de foco, é maior do que um limite inferior e menor do que um limite superior, a função de janela de ganho é configurada para retornar um ganho de janela sendo maior do que qualquer ganho de janela retornado pela função de janela de ganho, se o argumento de função de janela é menor do que o limite inferior, ou maior do que o limite superior.

[317] Por exemplo, no caso da direção de foco, a direção de foco pode por si própria o argumento de função de janela (e, portanto, o argumento de função de janela depende da direção do foco). No caso da posição de foco, um argumento de função de janela, pode, por exemplo, ser derivado a partir da posição de foco.

[318] Da mesma forma, a invenção pode ser aplicada a outros dispositivos portáteis, que incluem dispositivos de auxílio a audição ou dispositivos, tais como o Google Glass®. Deve-se notar que alguns dispositivos de vestir também são equipados com uma ou mais câmaras ou sensor de ToF que podem ser usados para estimar a distância de objetos para a pessoa que usa o dispositivo.

[319] Embora alguns aspectos tenham sido descritos no contexto de um aparelho, é claro que estes aspectos também representam uma descrição do método correspondente, onde um bloco ou dispositivo corresponde a uma etapa do método, ou uma característica de uma etapa do método. Analogamente, os aspectos descritos no contexto de uma etapa do método também representam uma descrição de um bloco correspondente ou item ou característica de um aparelho correspondente.

[320] O sinal decomposto da invenção pode ser armazenado em um meio de armazenamento digital ou pode ser transmitido sobre um meio de transmissão, tal como um meio de transmissão sem fios ou um meio de transmissão com fios, tais como a Internet.

[321] Dependendo certos requisitos de aplicação, as modalidades da invenção podem ser implementadas em hardware ou em software. A aplicação pode ser realizada usando um meio de armazenamento digital, por exemplo, um

disquete, um DVD, um CD, um ROM, um PROM, um EPROM, um EEPROM ou uma memória flash, com sinais de controle eletronicamente legíveis armazenados nos mesmos, os quais cooperam (ou são capazes de cooperar) com um sistema de computador programável de modo que o respectivo método é realizado.

[322] Algumas modalidades de acordo com a invenção compreendem um suporte de dados não transitórios com sinais de controle legíveis eletronicamente, os quais são capazes de cooperar com um sistema de computador programável, de tal modo que um dos métodos aqui descritos seja realizado.

[323] Geralmente, as modalidades da presente invenção podem ser implementadas como um produto de programa de computador com um código de programa, o código de programa sendo operativo para a realização de um dos métodos em que o produto de programa de computador é executado em um computador. O código de programa pode, por exemplo, ser armazenado em um suporte legível por máquina.

[324] Outras modalidades compreendem o programa de computador para a execução de um dos métodos aqui descritos, armazenados em um suporte legível por máquina.

[325] Em outras palavras, uma modalidade do método da invenção é, portanto, um programa de computador tendo um código de programa para realizar um dos métodos aqui descritos, quando o programa de computador é executado em um computador.

[326] Uma outra modalidade dos métodos da invenção é, portanto, um suporte de dados (ou um meio de armazenamento digital, ou um meio legível por computador) que compreende, nele gravado, o programa de computador para a execução de um dos métodos aqui descritos.

[327] Uma modalidade adicional do método da invenção é, como consequência, um fluxo de dados ou uma sequência de sinais, que representam o programa de computador para a execução de 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 conexão de comunicação de dados, por exemplo, através da Internet.

[328] Uma outra modalidade compreende um meio de processamento, por exemplo, um computador, ou um dispositivo lógico programável, configurado ou adaptado para executar um dos métodos aqui descritos.

[329] Uma modalidade adicional compreende um computador tendo nele instalado o programa de computador para a execução de um dos métodos aqui descritos.

[330] Em algumas modalidades, um dispositivo de lógica programável (por exemplo, uma disposição de portas programáveis de campo) pode ser usado para realizar algumas ou todas as funcionalidades dos métodos aqui descritos. Em algumas modalidades, uma disposição de portas programáveis de campo pode cooperar com um microprocessador para executar um dos métodos aqui descritos. Geralmente, os métodos são, de preferência, realizados por qualquer aparelho de hardware.

[331] As modalidades acima descritas são meramente ilustrativas dos princípios da presente invenção. Entende-se que modificações e variações dos arranjos e detalhes aqui descritos serão evidentes para outros especializados na técnica. É intenção, portanto, se ater apenas ao escopo das reivindicações de patente iminentes e não aos detalhes específicos apresentados a título de descrição e explicação das modalidades aqui.

Referências:

[1] Y. Ishigaki, M. Yamamoto, K. Totsuka, e N. Miyaji, “Zoom microphone,” in Audio Engineering Society Convention 67, Paper 1713, outubro de 1980.

[2] M. Matsumoto, H. Naono, H. Saitoh, K. Fujimura, e Y. Yasuno, “Stereo zoom microphone for consumer video cameras,” Consumer Electronics, IEEE Transactions on, vol. 35, n°. 4, pp. 759–766, novembro de 1989. agosto de 13, 2014

[3] T. van Waterschoot, W. J. Tirry, e M. Moonen, “Acoustic zooming by multi microphone sound scene manipulation,” J. Audio Eng. Soc, vol. 61, n°. 7/8, pp. 489–507, 2013.

[4] V. Pulkki, “Spatial sound reproduction with diretional audio coding,” J. Audio Eng. Soc, vol. 55, no. 6, pp. 503–516, junho de 2007.

[5] R. Schultz-Amling, F. Kuech, O. Thiergart, e M. Kallinger,

“Acoustical zooming based on a parametric sound field representation,” in Audio Engineering Society Convention 128, Paper 8120, London UK, maio de 2010.

[6] O. Thiergart, G. Del Galdo, M. Taseska, e. Habets, “Geometry-based spatial sound acquisition using distributed microphone arrays,” Audio, Speech, and Language Processing, IEEE Transactions on, vol. 21, n°. 12, pp. 2583–2594, dezembro de 2013.

[7] K. Kowalczyk, O. Thiergart, A. Craciun, e. A. P. Habets, “Sound acquisition in noisy and reverberant environments using virtual microphones,” em Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013 IEEE Workshop on, outubro de 2013.

[8] O. Thiergart e. A. P. Habets, “An informed LCMV filter based on multiple instantaneous diretion-of-arrival estimates,” em Acoustics Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013, pp. 659–663.

[9] O. Thiergart e. A. P. Habets, “Extracting reverberant sound using a linearly constrained minimum variance spatial filter,” Signal Processing Letters, IEEE, vol. 21, n°. 5, pp. 630–634, maio de 2014.

[10] R. Roy e T. Kailath, “ESPRIT-estimation of signal parameters via rotational invariance techniques,” Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 37, n°. 7, pp. 984–995, julho de 1989.

[11] B. Rao e K. Hari, “Performance analysis of root-music,” em Signals, Systems and Computers, 1988. Twenty-Second Asilomar Conference on, vol. 2, 1988, pp. 578–582.

[12] H. Teutsch e G. Elko, “An adaptive close-talking microphone array,” in Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the, 2001, pp. 163–166.

[13] 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, n°. 4, pp. 2337–2346, 2012.

[14] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” J. Audio Eng. Soc, vol. 45, n°. 6, pp. 456–466, 1997.

[15] J. Blauert, Spatial hearing, 3° ed. Hirzel-Verlag, 2001.

[16] T. May, S. van de Par, e A. Kohlrausch, “A probabilistic model for robust localization based on a binaural auditory front-end,” IEEE Trans. Audio, Speech, Lang. Process., vol. 19, n°. 1, pp. 1–13, 2011.

[17] J. Ahonen, V. Sivonen, e V. Pulkki, “Parametric spatial sound processing applied to bilateral hearing aids,” em AES 45° International Conference, março de 2012.

REIVINDICAÇÕES

1. Sistema para gerar um ou mais sinais de saída de áudio, compreendendo:

um módulo de decomposição (101),

um processador de sinais (105), e

uma interface de saída (106),

caracterizado por o módulo de decomposição (101) ser configurado para receber dois ou mais sinais de entrada de áudio, em que o módulo de decomposição (101) é configurado para gerar um sinal de componente direto, que compreende componentes de sinal direto de dois ou mais sinais de entrada de áudio, e em que o módulo de decomposição (101) é configurado para gerar um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de entrada de áudio,

em que o processador de sinais (105) é configurado para receber o sinal de componente direto, o sinal de componente difuso e as informações de direção, as ditas informações de direção dependentes de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio,

em que o processador de sinais (105) é configurado para gerar um ou mais sinais processados difusos dependendo do sinal de componente difuso,

em que, para cada sinal de saída de áudio de um ou mais sinais de saída de áudio, o processador de sinais (105) é configurado para determinar, dependendo da direção de chegada, um ganho direto, o processador de sinais (105) é configurado para aplicar o dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e o processador de sinais (105) é configurado para combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio, e

em que a interface de saída (106) é configurada para emitir o um ou mais sinais de saída de áudio,

em que para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio de uma função de ganho de panning é atribuída ao dito sinal de

saída de áudio,

em que a função de ganho de panning de cada um de dois ou mais sinais de saída de áudio compreende uma pluralidade de valores de argumento de função de panning, em que um valor de retorno de função panning é atribuído a cada um dos ditos valores de argumento de função de panning, em que, quando a dita função de ganho de panning recebe um dos ditos valores de argumento de função de panning, dita função de ganho de panning é configurada para retornar o valor de retorno de função de panning sendo atribuído ao dito um dos ditos valores de argumento de função de panning, em que a função de ganho de panning compreende um valor de argumento dependente de direção que depende da direção de chegada,

em que o processador de sinais (105) compreende um módulo de computação de função de ganho (104) para computar uma função de ganho direto para cada um de dois ou mais sinais de saída de áudio dependendo de uma função de ganho de panning sendo atribuída ao dito sinal de saída de áudio, e dependendo de uma função de janela de ganho, para determinar o ganho do dito sinal de saída de áudio,

em que o processador de sinais (105) é configurado para receber, adicionalmente, informações de orientação indicando um desvio angular de uma direção de olhar em relação à câmara, e pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende das informações de orientação; ou em que o módulo de computação de função de ganho (104) é configurado para receber informações de zoom, e as informações de zoom indicam um ângulo de abertura da câmara, e em que pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende das informações de zoom.

2. Sistema de acordo com a reivindicação 1, **caracterizado por** a função de ganho de panning de cada um dos dois ou mais de sinais de saída de áudio ter um ou mais máximos globais, sendo um dos valores de argumento de função de panning, em que para cada um de um ou mais máximos globais de cada função de ganho de panning, nenhum outro valor de argumento de função de panning existe para o qual dita função de ganho de panning retorna um valor de retorno de função de panning maior do que para o dito máximo

global, e

em que, para cada par de um primeiro sinal de saída de áudio e um segundo sinal de saída de áudio de dois ou mais sinais de saída de áudio, pelo menos um de um ou mais máximos globais da função de ganho de panning do primeiro sinal de saída de áudio é diferente de qualquer um de um ou mais máximos globais da função de ganho de panning do segundo sinal de saída de áudio.

3. Sistema, de acordo com a reivindicação 1, **caracterizado por** o processador de sinais (105) ser configurado para gerar cada sinal de saída de áudio de dois ou mais sinais de saída de áudio dependendo de um ganho de função de janela,

em que a função de janela de ganho é configurada para retornar um valor de retorno de função de janela quando recebe um valor de argumento de função de janela,

em que, se o valor de argumento de função de janela é maior do que um limite de janela inferior e menor do que um limite de janela superior, a função de janela de ganho é configurada para retornar um valor de retorno de função de janela sendo maior do que qualquer valor de retorno de função de janela retornado pela função de janela de ganho, se o valor de argumento de função de janela for menor do que o limite inferior, ou maior do que o limite superior.

4. Sistema, de acordo com a reivindicação 1, **caracterizado por** o módulo de computação de função de ganho (104) ser configurado para receber, adicionalmente, um parâmetro de calibração, e em que pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende do parâmetro de calibração.

5. Sistema, de acordo com a reivindicação 1, **caracterizado por** o processador de sinais (105) ser configurado para receber as informações de distância,

em que o processador de sinais (105) é configurado para gerar cada sinal de saída de áudio de dois ou mais sinais de saída de áudio dependendo das informações de distância.

6. Sistema, de acordo com a reivindicação 5, **caracterizado por** o

processador de sinais (105) ser configurado para receber um valor de ângulo original dependendo de uma direção original de chegada, sendo a direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio, e ser configurado para receber as informações de distância,

em que o processador de sinais (105) é configurado para calcular um valor de ângulo modificado dependendo do valor de ângulo original e dependendo das informações de distância, e

em que o processador de sinais (105) é configurado para gerar cada sinal de saída de áudio de dois ou mais sinais de saída de áudio dependendo do valor de ângulo modificado.

7. Sistema, de acordo com a reivindicação 5, **caracterizado por** o processador de sinais (105) ser configurado para gerar os dois ou mais sinais de saída de áudio através da condução de filtragem passa-baixo, ou através da adição de som direto retardado, ou através da condução de atenuação de som direto, ou através da condução de alisamento temporal, ou através da condução de dispersão de chegada, ou através da condução de descorrelação.

8. Sistema, de acordo com a reivindicação 1, **caracterizado por** o processador de sinais (105) ser configurado para gerar dois ou mais canais de saída de áudio,

em que o processador de sinais (105) é configurado para aplicar um ganho difuso sobre o sinal de componente difuso para obter um sinal difuso intermediário, e

em que o processador de sinais (105) é configurado para gerar um ou mais sinais descorrelacionados a partir do sinal difuso intermediário através da condução de descorrelação,

em que o um ou mais sinais descorrelacionados formam o um ou mais sinais difusos processados, ou em que o sinal difuso intermediário e o um ou mais sinais descorrelacionados formam o um ou mais sinais difusos processados.

9. Sistema, de acordo com a reivindicação 1, **caracterizado por** o sinal de componente direto e um ou mais sinais de componente diretos adicionais formar um grupo de dois ou mais sinais de componente diretos, em que o módulo de decomposição (101) é configurado para gerar o um ou mais

sinais de componente diretos, compreendendo adicionalmente, componentes de sinal direto de dois ou mais sinais de entrada de áudio,

em que a direção de chegada e uma ou mais direções de chegadas adicionais formam um grupo de duas ou mais direções de chegadas, em que cada direção de chegada do grupo de duas ou mais direções de chegadas é atribuída a exatamente um sinal de componente direto do grupo de dois ou mais sinais de componente diretos, em que o número de sinais de componente diretos de dois ou mais sinais de componente diretos e o número de direção de chegadas de duas direções de chegadas é igual,

em que o processador de sinais (105) é configurado para receber o grupo dos dois ou mais sinais de componentes diretos, e o grupo de duas ou mais direções de chegadas, e

em que, para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio,

o processador de sinais (105) é configurado para determinar, para cada sinal de componente direto do grupo dos dois ou mais sinais de componente diretos, um ganho direto, dependendo da direção de chegada do dito sinal de componente direto,

o processador de sinais (105) é configurado para gerar um grupo de dois ou mais sinais diretos processados através da aplicação, para cada sinal de componente direto do grupo de dois ou mais sinais de componente diretos, o ganho direto do dito sinal de componente direto sobre o dito sinal de componente direto, e

o processador de sinais (105) é configurado para combinar um de um ou mais sinais difusos processados e cada sinal processado do grupo de dois ou mais sinais processados para gerar o dito sinal de saída de áudio.

10. Sistema, de acordo com a reivindicação 9, **caracterizado por** o número de sinais de componente diretos do grupo de dois ou mais sinais de componente diretos mais 1, é menor do que o número de sinais de entrada de áudio sendo recebido por uma interface de recepção (101) do sistema.

11. Aparelho auditivo ou dispositivo de audição auxiliar, **caracterizado por** compreender um sistema de acordo com a reivindicação 1.

12. Aparelho para a gerar dois ou mais sinais de saída de áudio,

compreendendo:

um processador de sinais (105), e

uma interface de saída (106),

caracterizado por o processador de sinais (105) ser configurado para receber um sinal de componente direto, compreendendo componentes de sinal direto de dois ou mais sinais de áudio originais, em que o processador de sinais (105) é configurado para receber um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de áudio original, e em que o processador de sinais (105) é configurado para receber as informações de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio,

em que o processador de sinais (105) é configurado para gerar um ou mais sinais difusos processados dependendo do sinal de componente difuso,

em que, para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio, o processador de sinais (105) é configurado para determinar, dependendo da direção de chegada, um ganho direto, o processador de sinais (105) é configurado para aplicar o dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e o processador de sinais (105) é configurado para combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio, e

em que a interface de saída (106) é configurada para produzir dois ou mais sinais de saída de áudio,

em que para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio, uma função de ganho de panning é atribuída ao dito sinal de saída de áudio,

em que a função de ganho de panning de cada um de dois ou mais sinais de saída de áudio compreende uma pluralidade de valores de argumento de função de panning, em que um valor de retorno de função panning é atribuído a cada um dos ditos valores de argumento de função de panning, em que, quando dita função de ganho de panning recebe um dos ditos valores de

argumento de função de panning, a dita função de ganho de panning é configurada para retornar o valor de retorno de função de panning sendo atribuído ao dito um dos ditos valores de argumento de função de panning, em que a função de ganho de filtração compreende um valor de argumento dependente de direção que depende da direção de chegada,

em que o processador de sinais (105) compreende um módulo de computação de função ganho (104) para computar uma função de ganho direto para cada um de dois ou mais sinais de saída de áudio, dependendo da função de ganho de panning sendo atribuída ao dito sinal de saída de áudio e dependendo de uma função de janela de ganho, para determinar o ganho direto do dito sinal de saída de áudio, e

em que o processador de sinais (105) é configurado para receber, adicionalmente, informações de orientação, indicando um desvio angular de uma direção de olhar de uma câmara, e pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende das informações de orientação; ou em que o módulo de computação de função de ganho (104) é configurado para receber, adicionalmente, informações de zoom, e as informações de zoom indicam um ângulo de abertura da câmara, e em que pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende das informações de zoom.

13. Método para gerar dois ou mais sinais de saída de áudio, **caracterizado por** compreender:

receber dois ou mais sinais de entrada de áudio,

gerar um sinal de componente direto, compreendendo componentes de sinal direto de dois ou mais sinais de entrada de áudio,

gerar um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de entrada de áudio,

receber as informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio,

gerar um ou mais sinais difusos transformados dependendo do sinal de componente difuso,

para cada sinal de saída de áudio de dois ou mais sinais de saída de

áudio, determinar, dependendo da direção de chegada, um ganho direto, aplicar o dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio, e

produzir os dois ou mais sinais de saída de áudio,

em que para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio, uma função de ganho de panning é atribuída ao dito sinal de saída de áudio, em que a função de ganho de panning de cada um de dois ou mais sinais de saída de áudio compreende uma pluralidade de valores de argumento de função de panning, em que um valor de retorno de função panning é atribuído a cada um dos ditos valores de argumento de função de panning, em que, quando dita função de ganho de panning recebe um dos ditos valores de argumento de função de panning, a dita função de ganho de panning é configurada para retornar o valor de retorno de função de panning sendo atribuído ao dito um dos ditos valores de argumento de função de panning, em que a função de ganho de panning compreende um valor de argumento dependente de direção que depende da direção de chegada,

em que o método compreende, adicionalmente, computar uma função de ganho direto para cada um de dois ou mais sinais de saída de áudio, dependendo da função de ganho de panning sendo atribuída ao dito sinal de saída de áudio e dependendo de uma função de janela de ganho, para determinar o ganho direto do dito sinal de saída de áudio, e

em que o método compreende, adicionalmente, receber informações de orientação, indicando um desvio angular de uma direção de olhar de uma câmara, e pelo menos uma dentre a função de ganho de panning e a função de ganho janela depende das informações de orientação; ou em que o método compreende, adicionalmente, receber informações de zoom, em que as informações de zoom indicam um ângulo de abertura da câmara, e em que pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende das informações de zoom.

14. Método para gerar dois ou mais sinais de saída de áudio, **caracterizado por** compreender:

receber um sinal de componente direto, que compreende

componentes de sinal direto de dois ou mais sinais de áudio originais,

receber um sinal de componente difuso, que compreende componentes de sinal difuso de dois ou mais sinais de áudio originais,

receber informações de direção, ditas informações de direção dependendo de uma direção de chegada dos componentes de sinal direto de dois ou mais sinais de entrada de áudio,

gerar um ou mais sinais difusos processados de acordo com o sinal de componente difuso,

para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio, determinar, dependendo da direção de chegada, um ganho direto, aplicar o dito ganho direto sobre o sinal de componente direto para obter um sinal direto processado, e combinar o dito sinal direto processado e um de um ou mais sinais difusos processados para gerar o dito sinal de saída de áudio, e

produzir os dois ou mais sinais de saída de áudio,

em que para cada sinal de saída de áudio de dois ou mais sinais de saída de áudio, uma função de ganho de panning é atribuída ao dito sinal de saída de áudio,

em que a função de ganho de panning de cada um dos dois ou mais sinais de saída de áudio compreende uma pluralidade de valores de argumento de função de panning, em que um valor de retorno de função panning é atribuído a cada um dos ditos valores de argumento de função de panning, em que, quando a dita função de ganho de panning recebe um dos ditos valores de argumento de função de panning, a dita função de ganho de panning é configurada para retornar o valor de retorno de função de panning sendo atribuído ao dito um dos ditos valores de argumento de função de panning, em que a função de ganho de filtração compreende um valor de argumento dependente de direção que depende da direção de chegada,

em que o método compreende, adicionalmente, computar uma função de ganho direto para cada um de dois ou mais sinais de saída de áudio, dependendo da função de ganho de panning sendo atribuída ao dito sinal de saída de áudio e dependendo de uma função de janela de ganho, para determinar o ganho direto do dito sinal de saída de áudio, e

em que o método compreende, adicionalmente, receber informações

de orientação indicando um desvio angular de uma direção de olhar de uma câmara, e pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende da informação de orientação; ou

em que o método compreende, adicionalmente, receber informações de zoom, em que as informações de zoom indicam um ângulo de abertura da câmara, e em que pelo menos uma dentre a função de ganho de panning e a função de janela de ganho depende das informações de zoom.

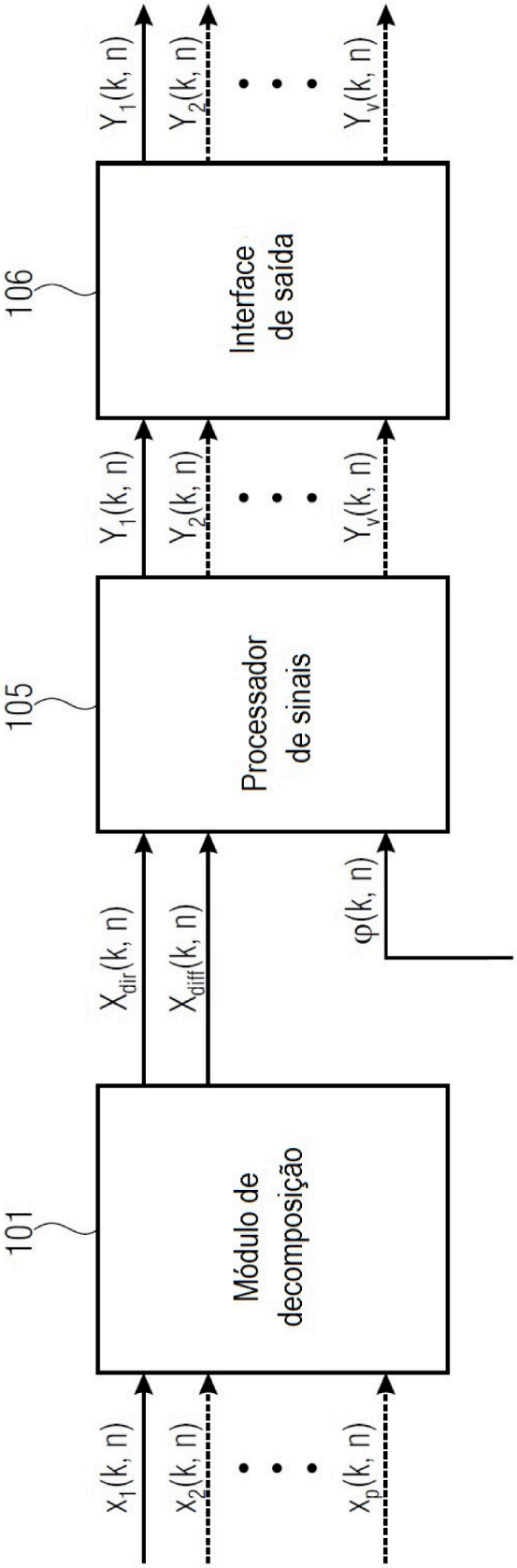

101
Módulo de decomposição
$x_1(k, n)$
$x_2(k, n)$
$x_p(k, n)$
$X_{dir}(k, n)$
$X_{diff}(k, n)$
105
Processador de sinais
$\varphi(k, n)$
$Y_1(k, n)$
$Y_2(k, n)$
$Y_v(k, n)$
106
Interface de saída
$Y_1(k, n)$
$Y_2(k, n)$
$Y_v(k, n)$

FIGURA 1A

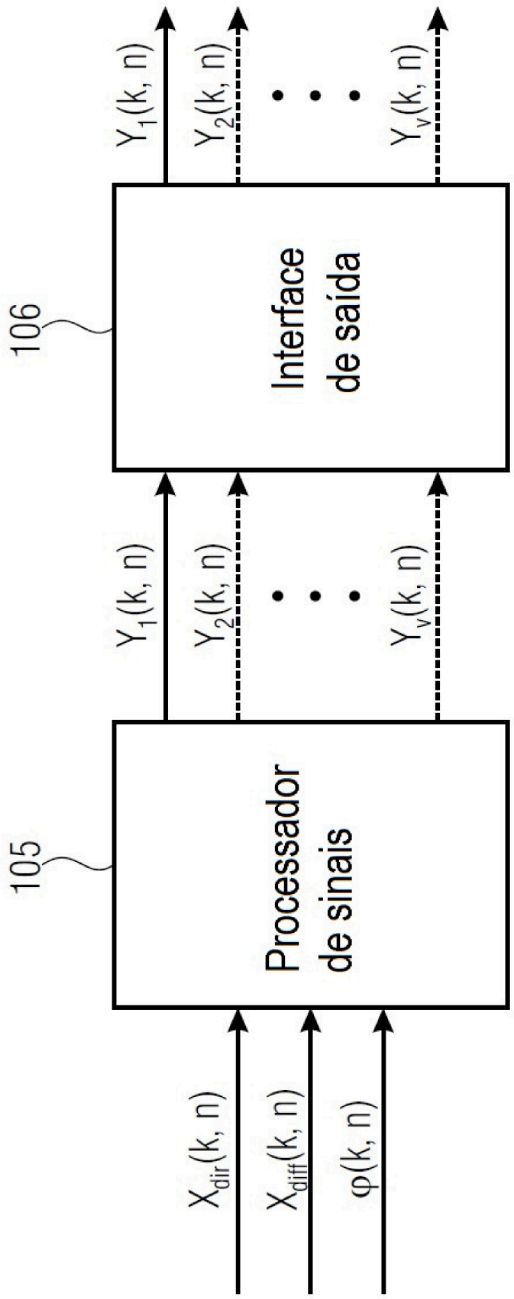

105
Processador de sinais
$X_{dir}(k, n)$
$X_{diff}(k, n)$
$\varphi(k, n)$
$Y_1(k, n)$
$Y_2(k, n)$
$Y_v(k, n)$
106
Interface de saída
$Y_1(k, n)$
$Y_2(k, n)$
$Y_v(k, n)$

FIGURA 1B

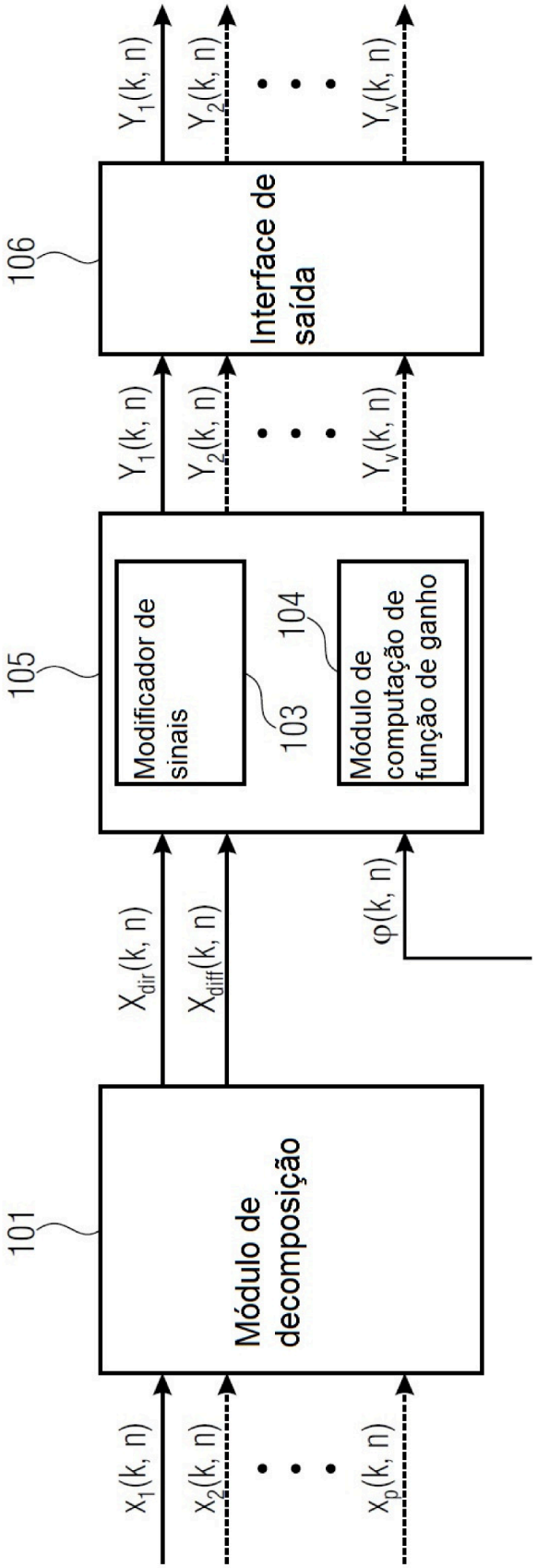

101
Módulo de decomposição
$x_1(k, n)$
$x_2(k, n)$
$x_p(k, n)$
$X_{dir}(k, n)$
$X_{diff}(k, n)$
$\varphi(k, n)$
105
Modificador de sinais
103
104
Módulo de computação de função de ganho
$Y_1(k, n)$
$Y_2(k, n)$
$Y_v(k, n)$
106
Interface de saída
$Y_1(k, n)$
$Y_2(k, n)$
$Y_v(k, n)$

FIGURA 1C

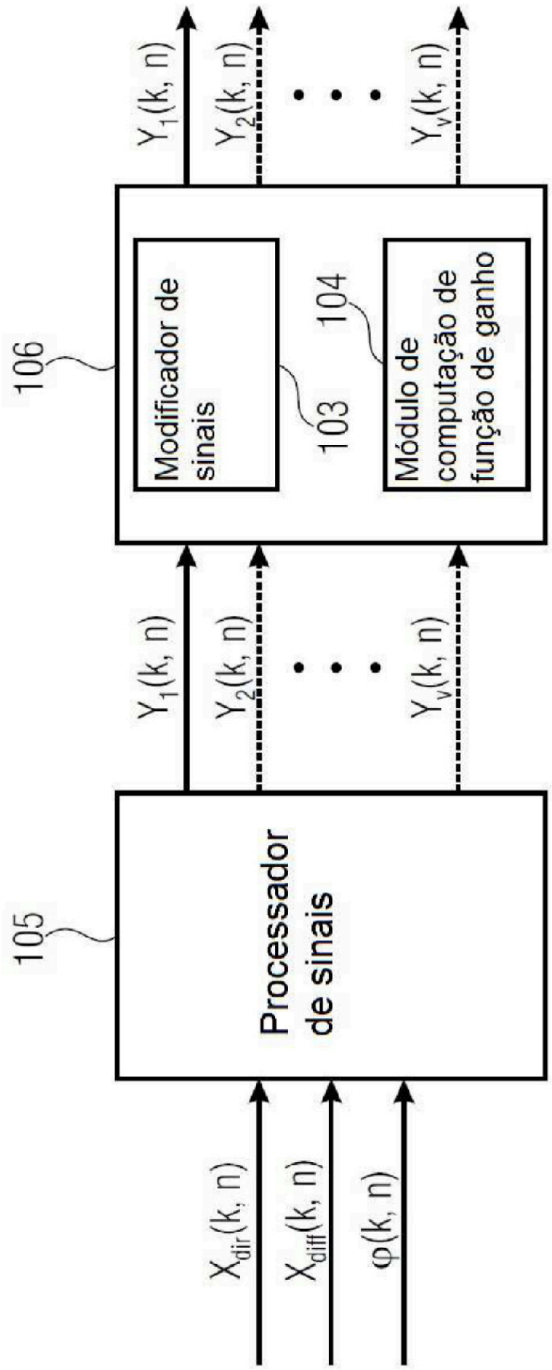

105
Processador de sinais
$X_{dir}(k, n)$
$X_{diff}(k, n)$
$\varphi(k, n)$
$Y_1(k, n)$
$Y_2(k, n)$
$Y_v(k, n)$
106
Modificador de sinais
103
104
Módulo de computação de função de ganho
$Y_1(k, n)$
$Y_2(k, n)$
$Y_v(k, n)$

FIGURA 1D

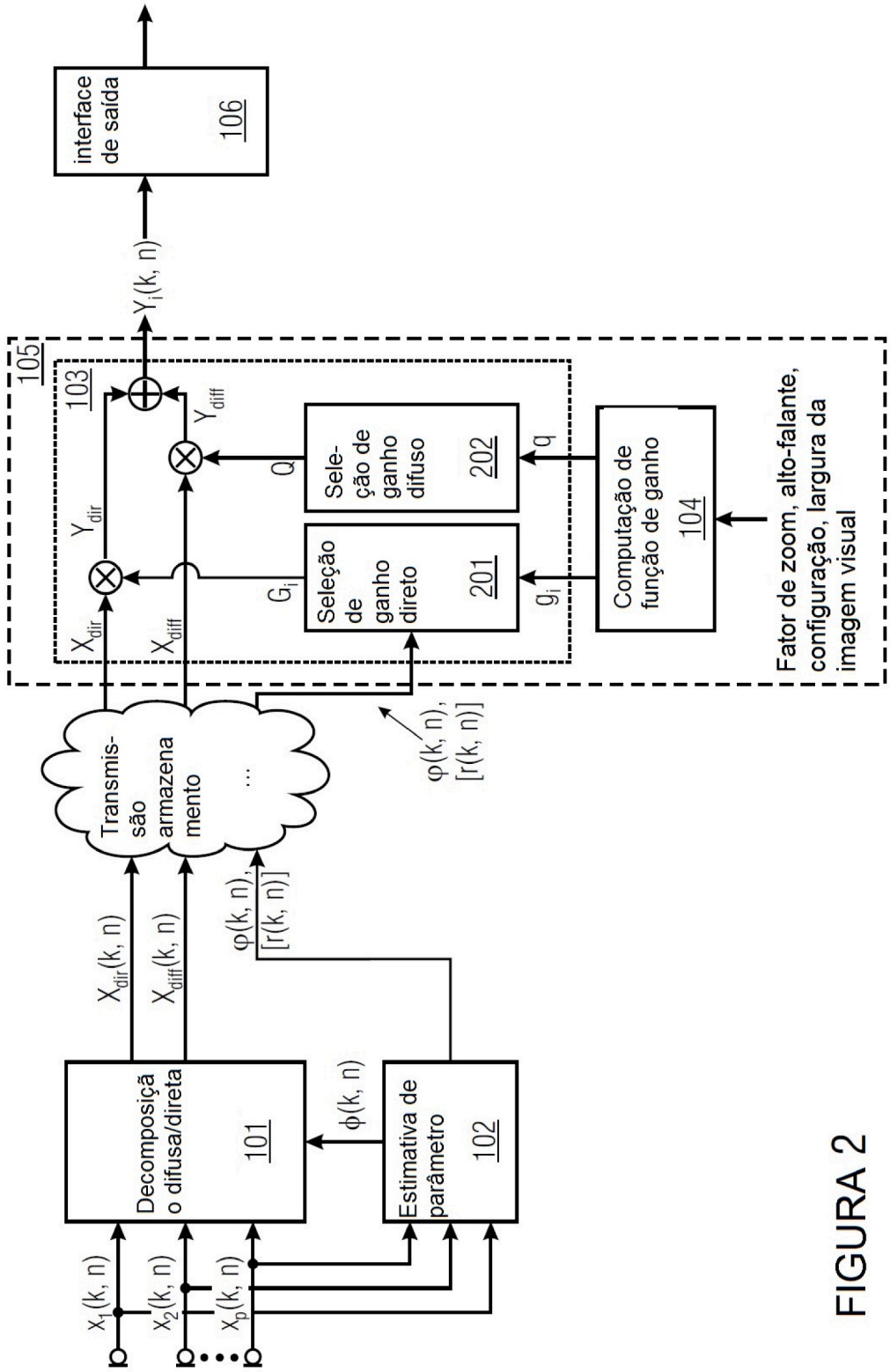

interface de saída
106
$Y_i(k, n)$
105
103
Y_{dir}
Y_{diff}
X_{dir}
X_{diff}
G_i
Q
Seleção de ganho direto
201
Sele-ção de ganho difuso
202
g_i
q
Computação de função de ganho
104
Fator de zoom, alto-falante, configuração, largura da imagem visual
Transmis-são armazena mento
...
$\varphi(k, n)$, $[r(k, n)]$
$X_{dir}(k, n)$
$X_{diff}(k, n)$
$\varphi(k, n)$, $[r(k, n)]$
Decomposição difusa/direta
101
$\phi(k, n)$
Estimativa de parâmetro
102
$x_1(k, n)$
$x_2(k, n)$
$x_p(k, n)$

FIGURA 2

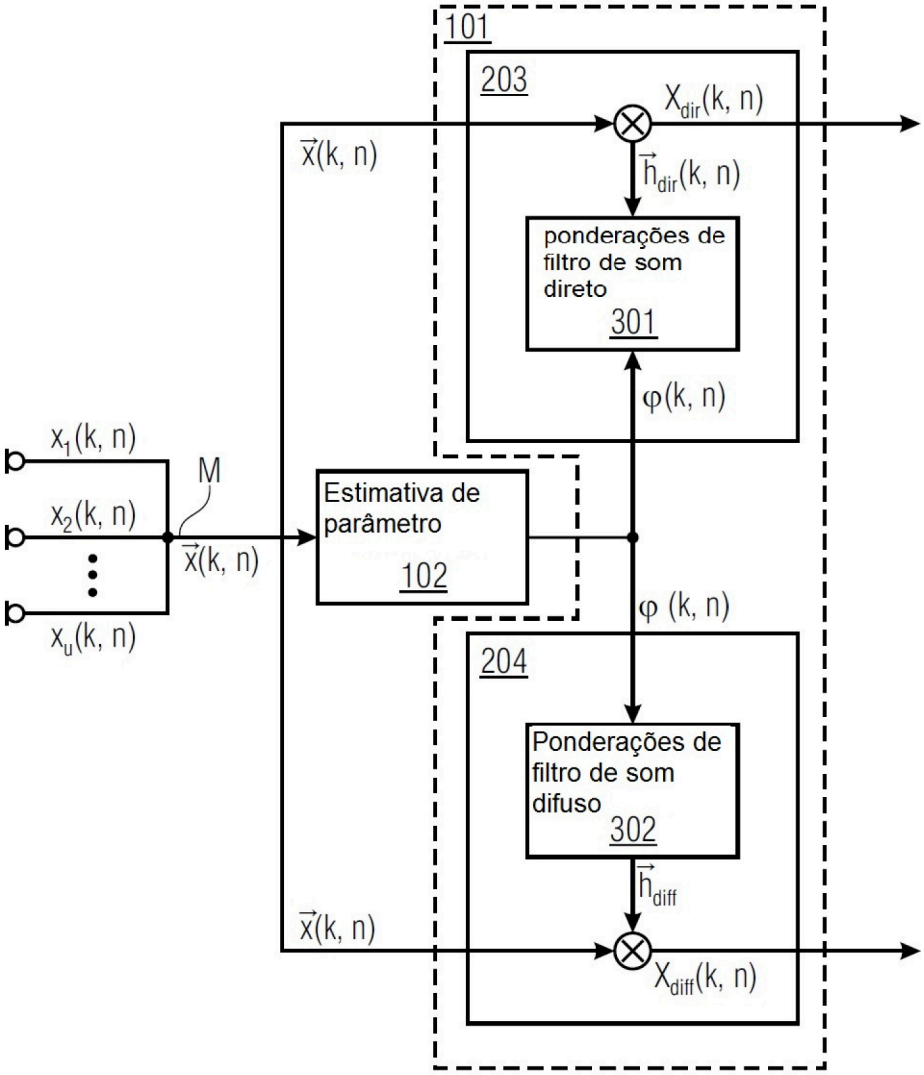

101
203
$X_{dir}(k, n)$
$\vec{x}(k, n)$
$\vec{h}_{dir}(k, n)$
ponderações de filtro de som direto
301
$\varphi(k, n)$
$x_1(k, n)$
M
$x_2(k, n)$
$\vec{x}(k, n)$
Estimativa de parâmetro
102
$x_u(k, n)$
φ (k, n)
204
Ponderações de filtro de som difuso
302
$\vec{h}_{diff}$
$\vec{x}(k, n)$
$X_{diff}(k, n)$

FIGURA 3

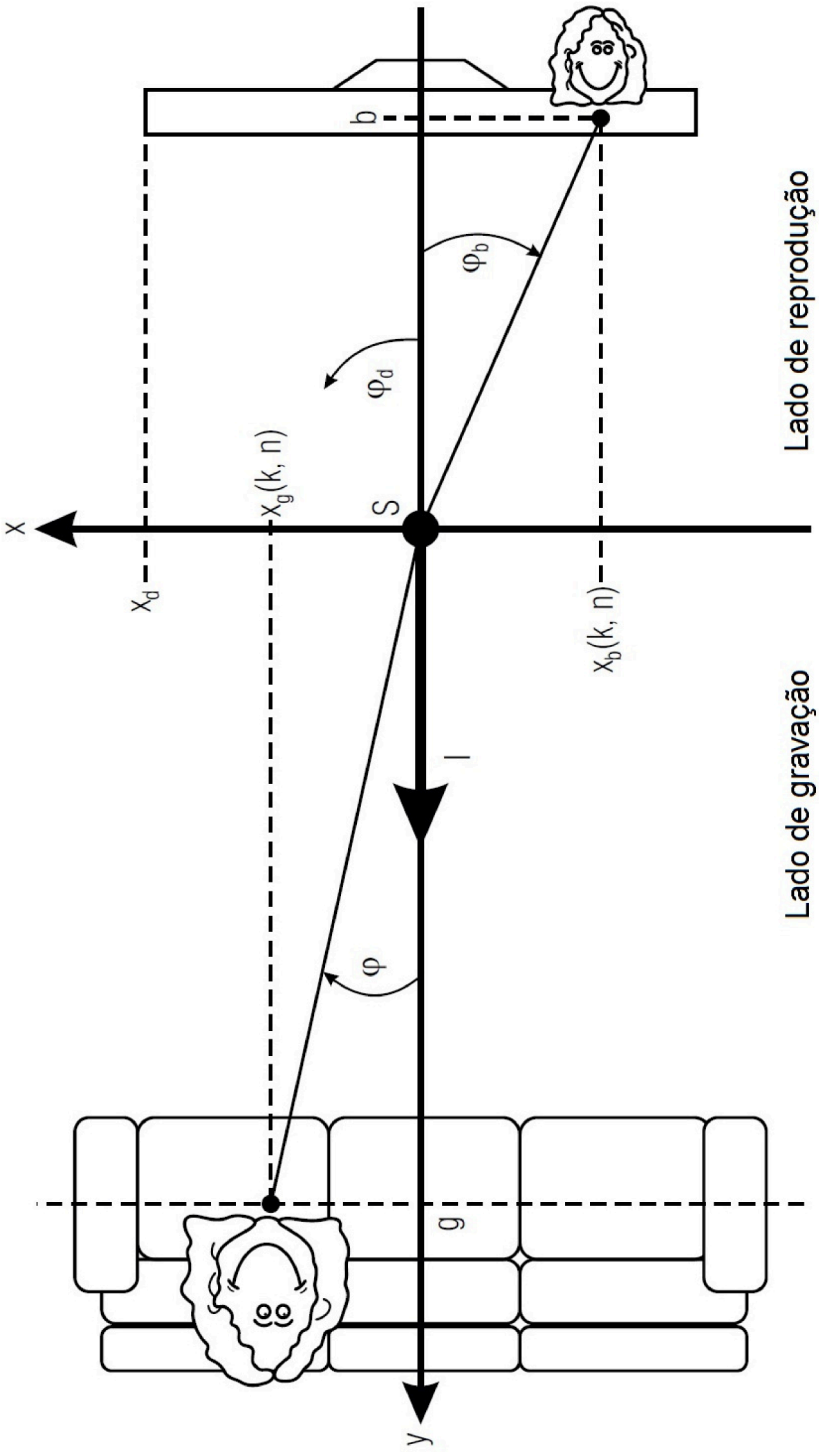
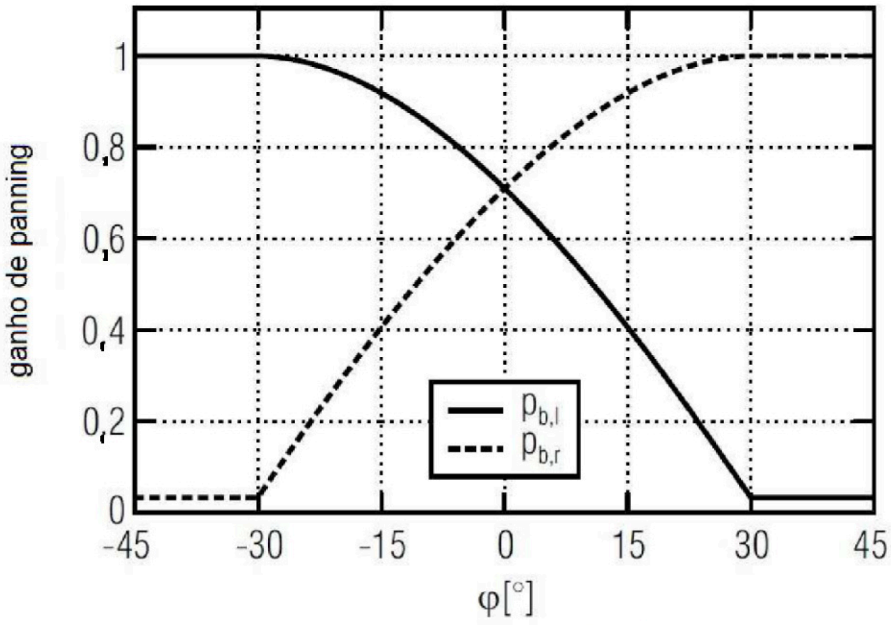
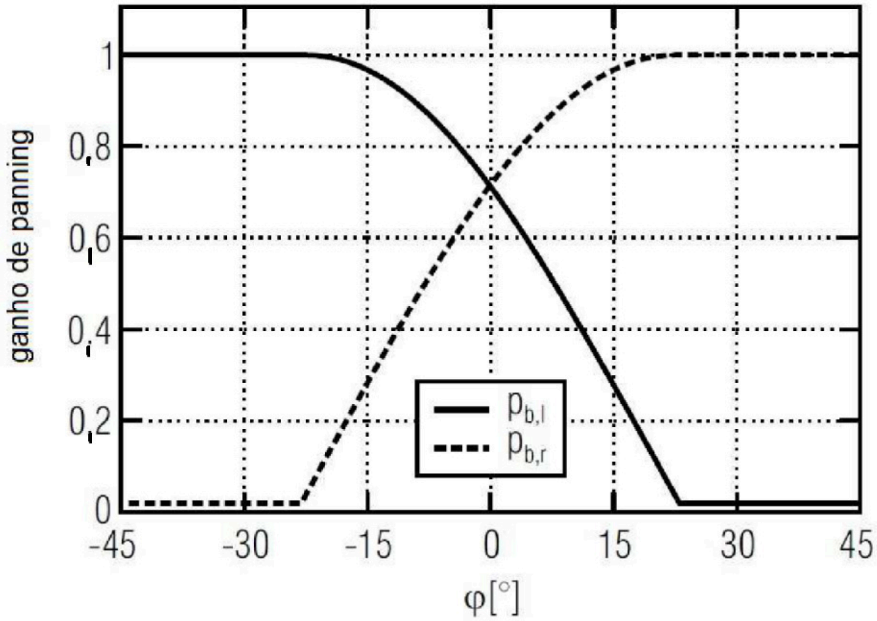

x
y
x_d
$x_g(k, n)$
S
φ_d
φ_b
b
$x_b(k, n)$
l
φ
g
Lado de reprodução
Lado de gravação

FIGURA 4



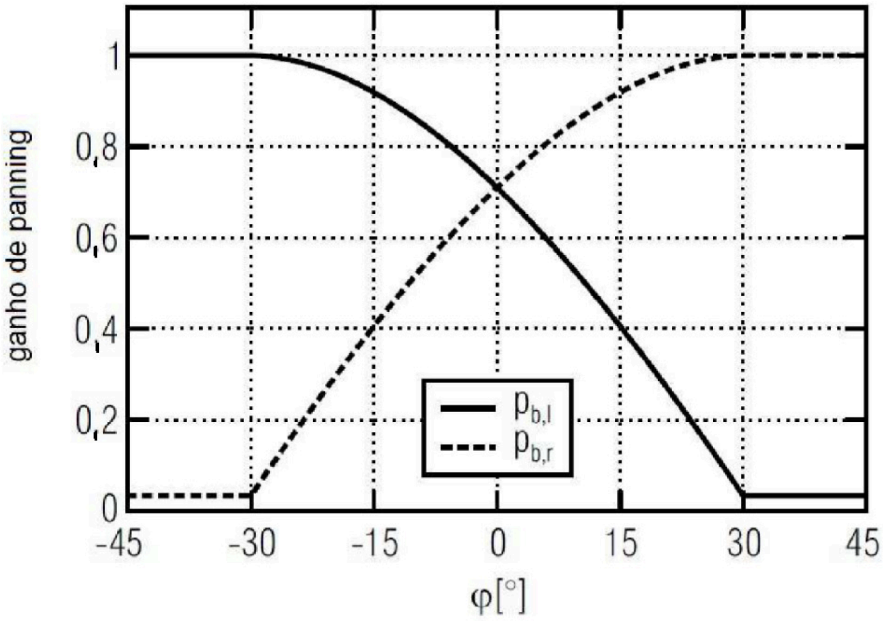
Função de panning VBAP

FIGURA 5A



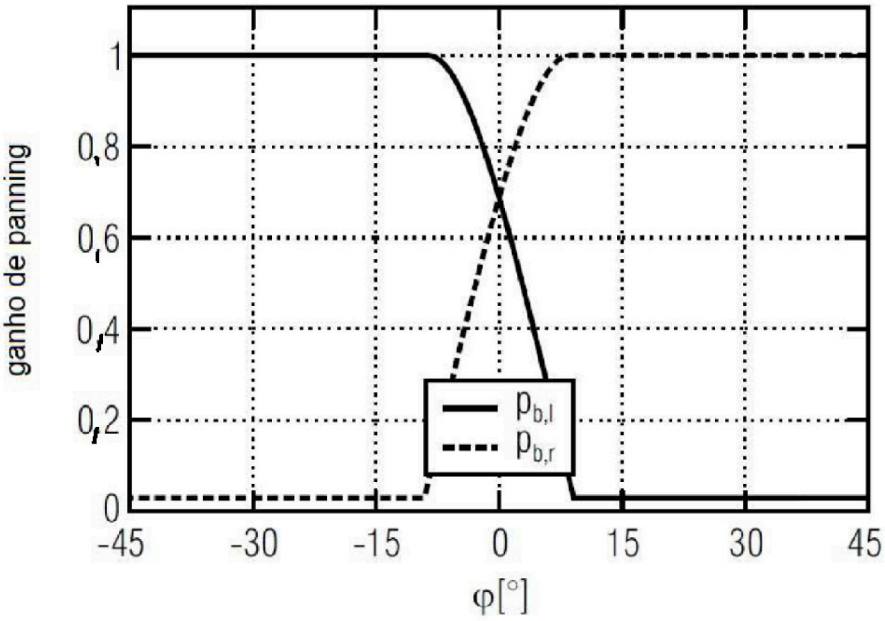
Função de panning para reprodução consistente

FIGURA 5B



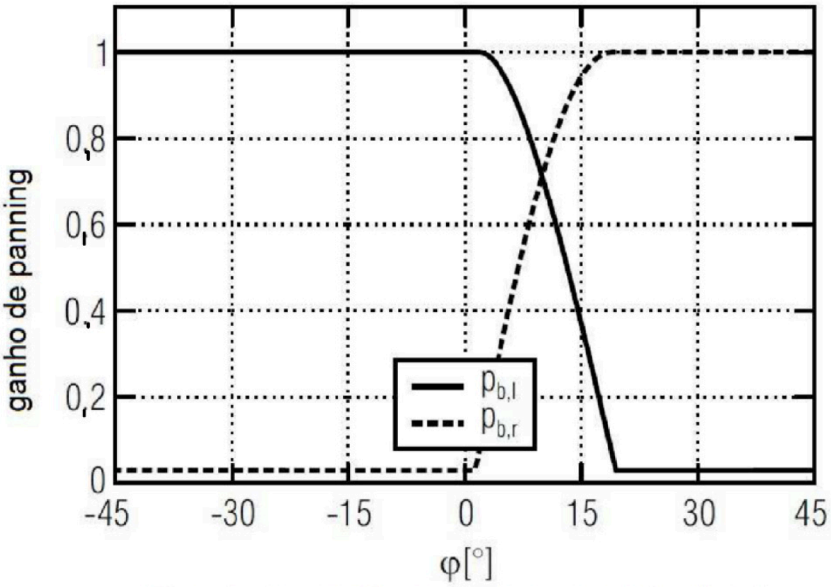
função de panning VBAP

FIGURA 6A



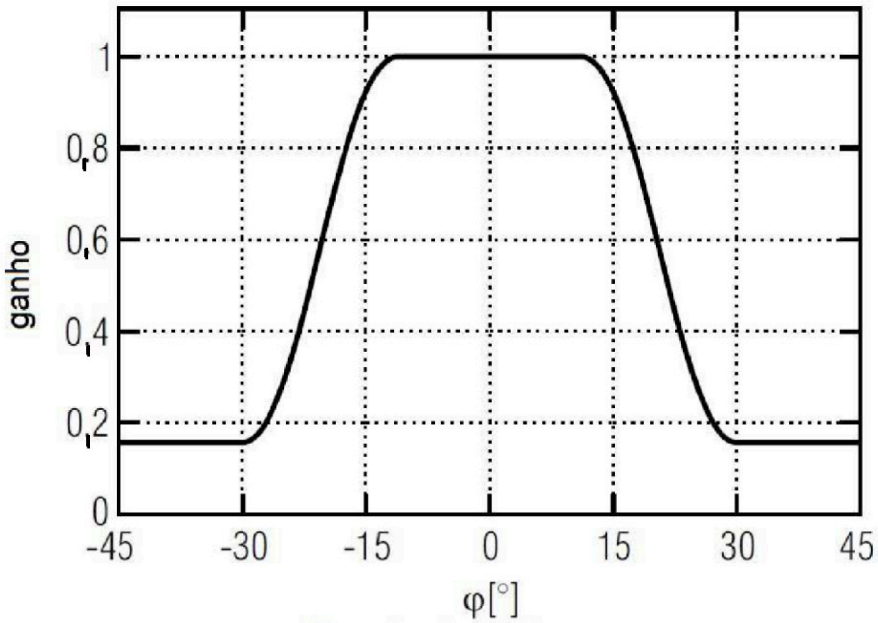
Função de panning após zoom

FIGURA 6B



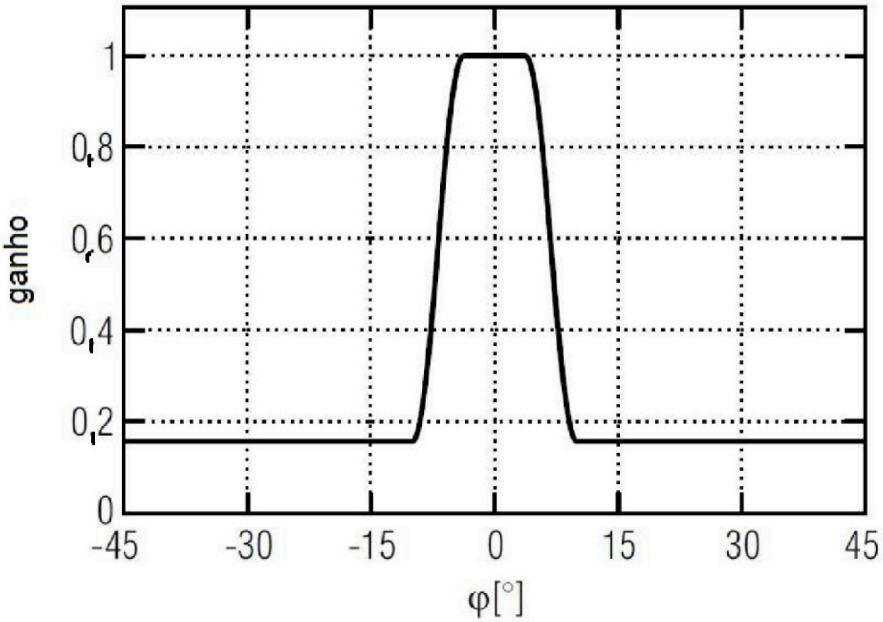
Função de panning após zoom com um desvio

FIGURA 6C



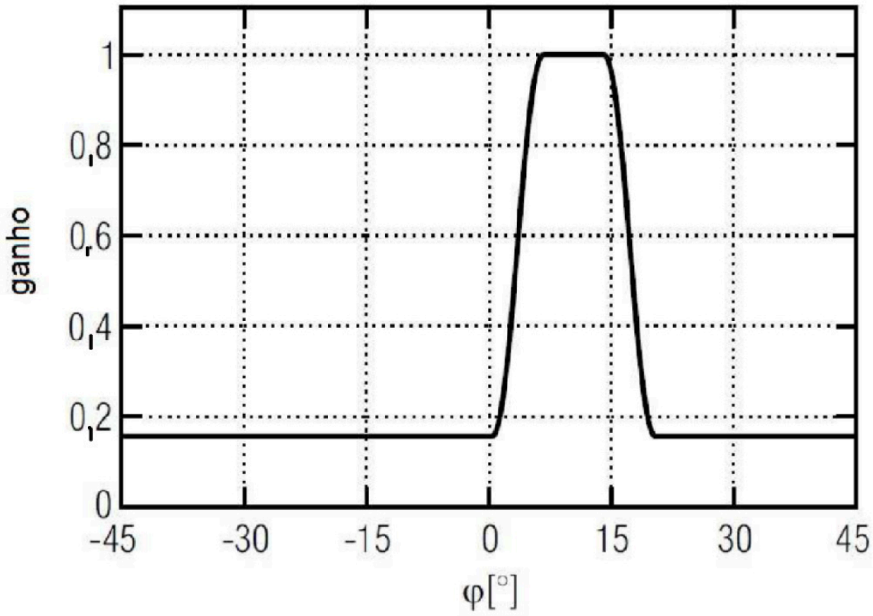
Função de janela

FIGURA 7A



Função de janela após zoom

FIGURA 7B



Função de janela após zoom com um desvio

FIGURA 7C

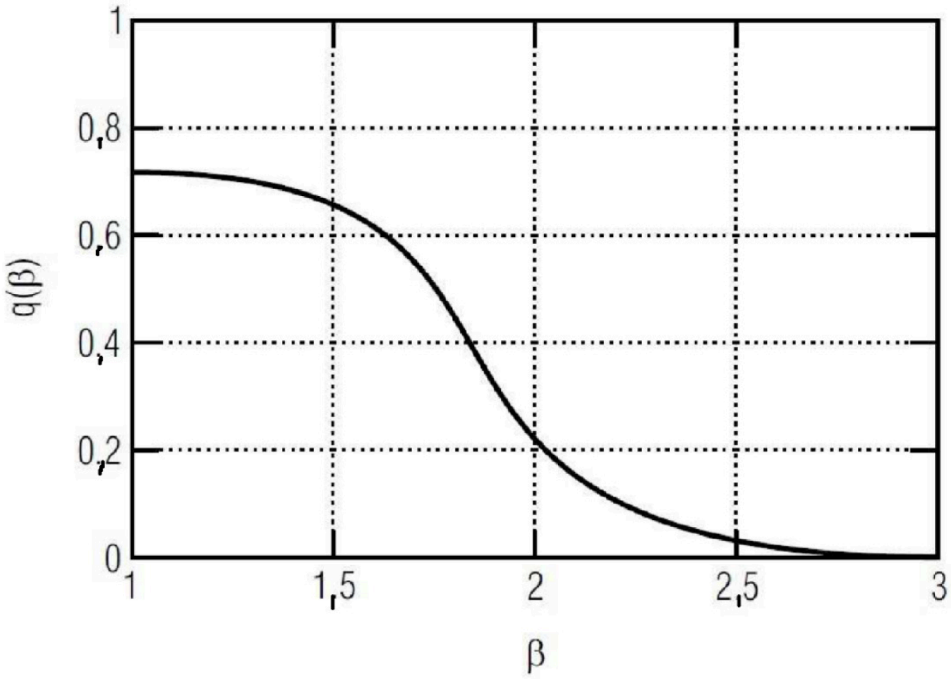


FIGURA 8

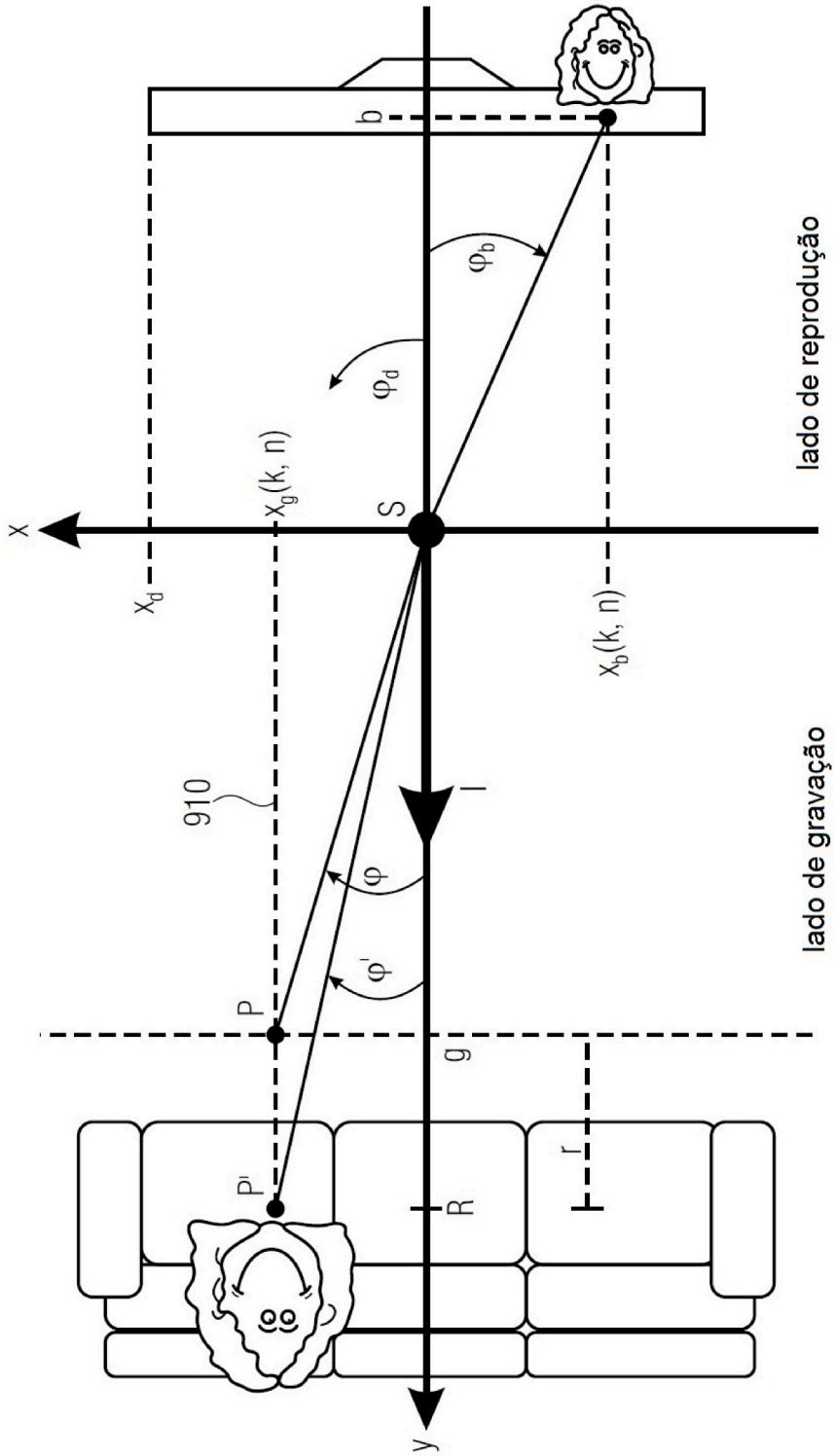
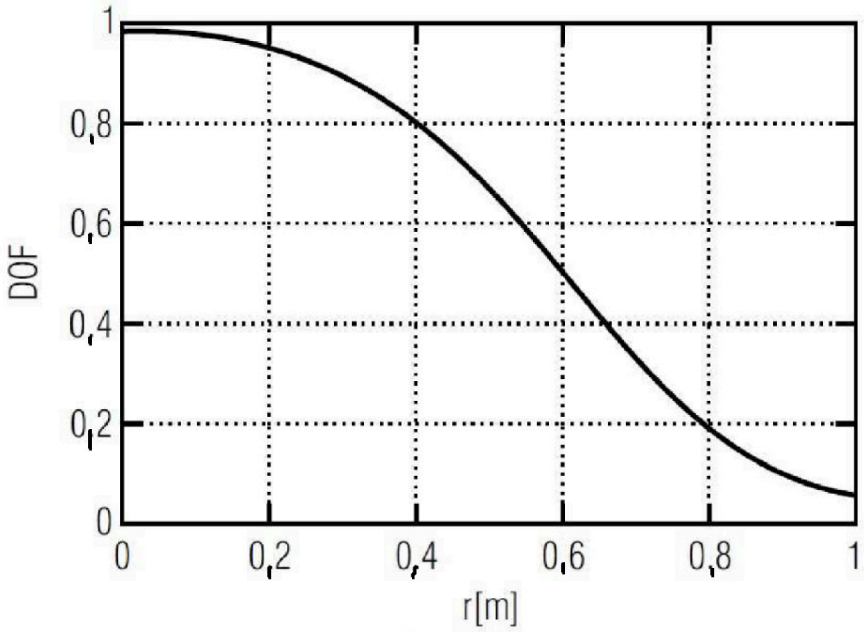
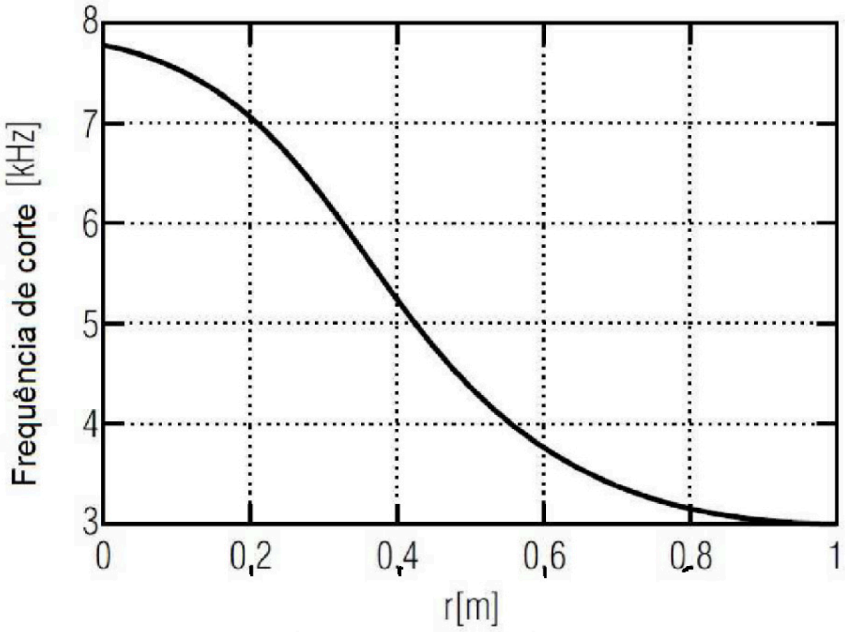

x
y
x_d
$x_g(k, n)$
S
$x_b(k, n)$
b
φ_d
φ_b
910
P
P'
φ
φ'
l
g
R
r
lado de gravação
lado de reprodução

FIGURA 9



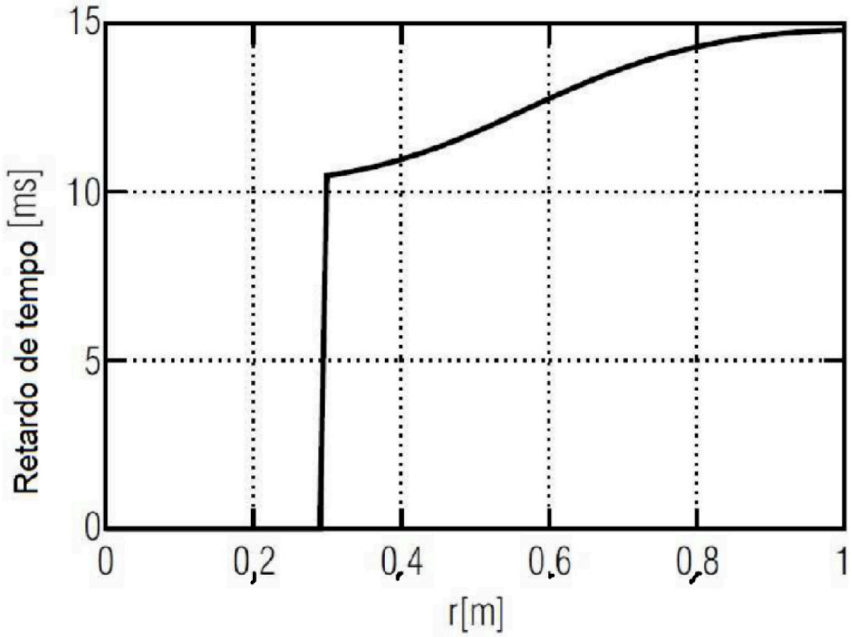
profundidade do campo

FIGURA10A



Frequência de corte de passa-baixa

FIGURA 10B



Retardo de tempo do som direto repetido

FIGURA 10C

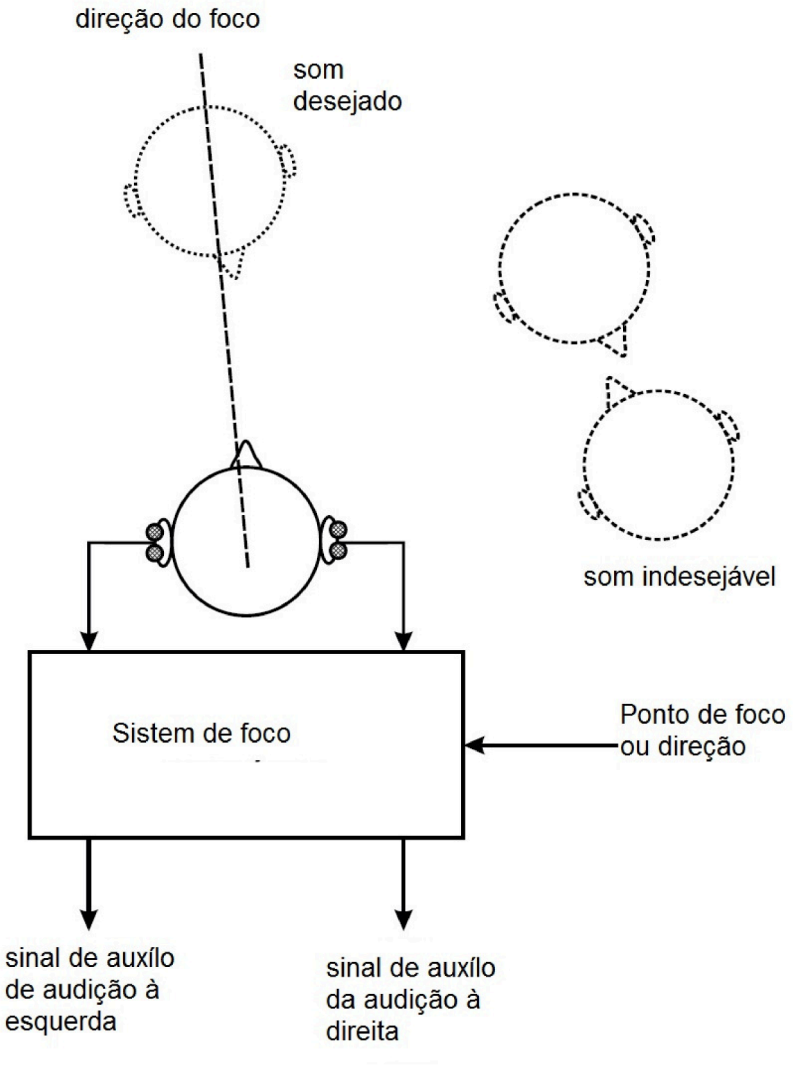

direção do foco
som
desejado
som indesejável
Sistem de foco
Ponto de foco
ou direção
sinal de auxílo
de audição à
esquerda
sinal de auxílo
da audição à
direita

FIGURA 11