



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

(11) PI 0316548-5 B1

(22) Data do Depósito: 28/11/2003

(45) Data de Concessão: 27/12/2016



(54) Título: MÉTODO PARA DESCRIÇÃO DE COMPOSIÇÃO DE SINAIS DE ÁUDIO

(51) Int.Cl.: H04S 3/00

(30) Prioridade Unionista: 02/12/2002 EP 02026770.4, 15/07/2003 EP 03016029.5

(73) Titular(es): THOMSON LICENSING S.A.

(72) Inventor(es): JENS SPILLE; JÜRGEN SCHMIDT

"MÉTODO PARA DESCRIÇÃO DE COMPOSIÇÃO DE SINAIS DE ÁUDIO"

A invenção refere-se a um método e a um dispositivo para codificação e decodificação de uma descrição de apresentação de sinais de áudio, especialmente para o posicionamento no espaço de sinais de áudio MPEG-4 codificados em um domínio de 3D.

Antecedentes

O padrão de áudio MPEG-4, como definido no padrão de áudio MPEG-4 ISO/IEC 14496 - 3 : 2001 e no padrão MPEG-4 14496 - 1 : 2001 do sistema MPEG-4, facilita uma ampla variedade de aplicações por meio de suporte da representação de objetos de áudio. Para a combinação dos objetos de áudio, informações adicionais, as assim denominadas descrições de cenário - determinam o posicionamento no espaço e no tempo sendo transmitido em conjunto com os objetos de áudio codificados.

Para reprodução os objetos de áudio são decodificados separadamente e compostos com utilização de uma descrição de cenário, a fim de preparar uma única trilha sonora, a qual é reproduzida para o ouvinte.

Para fins de eficiência, o padrão ISAO / IEC 14496 - 1:2001 do sistema MPEG-4 define uma maneira de codificar o descrição de cenário, em uma representação binária, assim denominado formato binário para a descrição de cenário (BIFS). De forma correspondente, os cenários de áudio são descritos com uso dos assim chamados AudioBIFS.

Uma descrição de cenário encontra-se estruturada

de forma hierárquica e pode ser representada como um gráfico, em que nós de ramificação do gráfico configuram os objetos em separado e os outros nós descrevem o processamento, por exemplo, o posicionamento, escala, efeitos. A aparência e o comportamento dos objetos em separado podem ser controlados com uso de parâmetros dentro dos nós de descrição de cenário.

Invenção

A invenção se baseia no reconhecimento do seguinte fato. A versão acima mencionada do padrão de áudio MPEG-4 define um nó denominado "sound", que permite o posicionamento no espaço dos sinais de áudio em um domínio 3D. Um nó adicional com o nome de "sound2D" apenas permite o posicionamento no espaço em uma tela 2D. o uso do nó "sound" em um aparelho reproduzidor gráfico 2D não é especificado devido a diferentes implementações das propriedades em um aparelho reproduzidor 2D e 3D. No entanto, a partir de aplicações de jogos, cinema e TV se sabe que faz sentido prover ao usuário final uma apresentação de "som em 3D" com posicionamento espacial completo, mesmo em caso da apresentação de vídeo ser limitada à pequena tela plana em frente. Isso não é possível com os nós definidos "sound" e "som 2D".

Assim sendo, um problema a ser solucionado pela invenção é a superação da desvantagem acima mencionada.

Em princípio, o método de codificação da invenção inclui a geração de uma descrição de parâmetros de uma fonte sonora, incluindo a informação, que permite o posicionamento no espaço em um sistema de coordenadas 2D. A descrição de

parâmetros da fonte sonora encontra-se ligada aos sinais de da referida fonte sonora. Um valor adicional em 1D é adicionado à referida descrição de parâmetros, que permite, em um contexto visual em 2D, um posicionamento no espaço da referida fonte sonora em um domínio 3D.

Fontes de som separadas podem ser codificadas como objetos de áudio separados, e a disposição das fontes de som em um cenário de som pode ser descrita por uma descrição de cenário apresentando primeiros nós correspondendo aos objetos de áudio separados e segundos nós descrevendo a apresentação dos objetos de áudio. Um campo de um segundo nó pode definir o posicionamento no espaço em 3D de uma fonte sonora.

Vantajosamente, o sistema de coordenadas 2D corresponde ao plano de tela e o valor em 1D corresponde a uma informação de profundidade, perpendicular ao referido plano de tela.

Além disso, uma transformação dos referidos valores do sistema de coordenadas em 2D para referidas posições em 3 dimensões pode possibilitar o movimento de um objeto gráfico em um plano de tela a ser mapeado em relação ao movimento de um objeto de áudio em uma profundidade perpendicular ao referido plano de tela.

O método de decodificação da invenção inclui, em princípio, a recepção de um sinal de áudio correspondendo a uma fonte sonora ligada à descrição de parâmetros da fonte sonora. A descrição de parâmetros inclui informação que permite o posicionamento espacial em um sistema de coordenadas

2D. um valor adicional em 1D é separado da descrição de parâmetros. A fonte sonora é posicionada no espaço em um contexto visual em 2D em um domínio em 3D, utilizando-se o referido valor adicional em 1D.

5 Objetos de áudio representando fontes de som separadas podem ser decodificados em separado e uma única trilha sonora pode ser composta a partir dos objetos de áudio decodificados, utilizando-se uma descrição de cenário possuindo primeiros nós, correspondendo aos objetos de áudio separados
10 e segundos nós, descrevendo o processamento dos objetos de áudio. Um campo de um segundo nó pode definir o posicionamento no espaço em 3D de uma fonte sonora.

Vantajosamente, o sistema de coordenadas 2D corresponde ao plano de tela, e o referido valor em 1D corresponde a uma informação de profundidade, perpendicular ao referido plano de tela.
15

Além disso, a transformação dos referidos valores do sistema de coordenadas 2D em referidas posições em 3 dimensões pode possibilitar o movimento de um objeto gráfico
20 no plano de cenário a ser mapeado em relação a um movimento de um objeto de áudio em uma profundidade perpendicular ao referido plano de cenário.

Exemplos de Modalidades

O nó sound2D se define como se segue:

```
25 Sound2D {
    exposedField    SFFFloat  Intensidade          1,0

    ExposedField    SFVec2f    Localização          0,0
```

```

ExposedField  SFNó      Fonte                      Nulo
ExposedField  SFBool     Posicionamento espacial Verdadeiro
}

```

e o nó sound, que se trata de um nó 3D, se define como se segue:

```

sound{
ExposedField  SFVec3f    Direção                      0, 0, 1

ExposedField  SFFloat    Intensidade                  1,0
ExposedField  SFVec3f    Localização                  0, 0, 0
ExposedField  SFFloat    MaxPosterior                 10.0
ExposedField  SFFloat    MaxFrontal                   10.0
ExposedField  SFFloat    MinPosterior                 1.0
ExposedField  SFFloat    MinFrontal                   1.0
ExposedField  SFFloat    Prioridade                   0.0
ExposedField  SFNode     Fonte                        Nulo
Field         SFBool     Posicionamento espacial Verdadeiro
5 }

```

A seguir, o termo geral para todos os nós de som (Sound2D, Sound e directiveSound) serão escritos com letra minúscula, por exemplo “nós de som”.

No caso mais simples, o nó Sound ou Sound2D fica
 10 conectado através de um nó de AudioSource à saída do decodificador. Os nós de som contêm a informação de intensidade e de localização.

A partir do ponto de vista do áudio, um nó de som trata-se do nó final, antes do mapeamento do alto-falante.
 15 No caso de diversos nós de som, a saída será somada. A par-

tir do ponto de vista dos sistemas, os nós de som podem ser visualizados como ponto de entrada para um sub-gráfico de áudio. Um nó de som pode ser agrupado com nós diferentes de áudio em um nó de transformação que irá ajustar seu local original.

Com o campo *phaseGroup* do nó *AudioSource*, é possível marcar canais que contenham relações de fase importantes, como no caso do "par estéreo", "multicanais" etc. É permitida uma operação mista de canais com fase relacionada e canais com fase não relacionada. Um campo *posicionamento espacial* nos nós de som especifica se o som deve ser posicionado no espaço ou não. Isso é apenas verdadeiro para canais, os quais não são membros de um grupo de fase.

O *sound2D* pode posicionar no espaço o som em uma tela em 2D. O padrão dita que o som deveria ser posicionado no espaço no cenário de dimensão 2m x 1,5m a uma distância de um metro. Esta explicação parece ser ineficaz devido ao valor do campo de localização não ser restrito, e portanto o som também pode ficar posicionado fora da dimensão da tela.

O nó *sound* e *DirectiveSound* pode ajustar a localização em qualquer lugar no espaço 3D. O mapeamento em relação ao posicionamento do alto-falante existente pode ser efetuado utilizando-se o panorâmico para a amplitude simples ou técnicas mais sofisticadas.

Tanto *sound*, quanto *sound2D* podem manipular entradas de canal múltiplo e basicamente apresentarem as mesmas funções, mas o nó *sound2D* não pode posicionar no espaço um som diferente de frontal.

Uma possibilidade é de adicionar sound e sound2D a todos os perfis de gráficos de cena, isto é, adicionar o nó sound ao grupo SF2DNode.

Contudo, uma razão para não inclusão dos nós de som "3D" nos perfis de gráfico de cenário 2D é que um típico aparelho reproduzidor em 2D não é capaz de manipular vetores em 3D (tipo SFVec3f), como seria necessário para os campos *direção* do som e *localização*.

Outra razão é que o nó sound encontra-se especialmente projetado para cenários de realidade virtual com pontos ouvintes em movimento e atributos de atenuação para objetos sonoros de grande distância. Por isso, o nó de ponto ouvinte e os campos maxPosterior, maxFrontal, minPosterior e minFrontal de som são definidos.

De acordo com uma modalidade, o nó sound2D antigo se estende ou um novo sound2DDepth é definido. O nó sound2DDepth poderia ser similar ao nó sound2D, mas com um campo de profundidade adicional.

```

Sound2DDepth{
  exposedField  SFFFloat  Intensidade          1,0

  ExposedField  SFVec2f   Localização           0,0
  ExposedField  SFFFloat  Profundidade          0,0
  ExposedField  SFNó      Fonte                 Nulo
  campo        SFBool    Posicionamento no espaço Verdadeiro
20 }

```

O campo *intensidade* ajusta o volume do som. Esta valor varia de 0,0 a 1,0 e este valor especifica um fator que é utilizado durante a reprodução do som.

O campo *localização* especifica a localização do som no cenário 2D.

O campo *profundidade* especifica a profundidade do som em um cenário 2D, utilizando o mesmo sistema de coordenadas do campo de localização. O valor padrão é 0,0 e se refere à posição da tela.

O campo *posicionamento espacial* especifica se o som deveria ser posicionado no espaço. Caso este indicador seja ajustado, o som deveria ser posicionado no espaço com o máximo de sofisticação possível.

As mesmas regras para o posicionamento espacial de áudio em múltiplos canais se aplicam ao nó Sound2DDepth, como ao nó de sound (3D).

A utilização do nó sound2D em um cenário 2D permite a apresentação de som envolvente, como o autor o registrou. Não é possível um *posicionamento no espaço* de um som diferente do frontal. Posicionamento espacial significa movimento do local de uma sinal mono fônico devido às interatividades de usuário ou atualizações de cenário.

Com o nó sound2DDepth é possível *posicionar no espaço* um som também de forma posterior, na lateral ou acima do ouvinte. Supondo-se que o sistema de apresentação de áudio tenha a capacidade de apresentá-lo.

A invenção não se restringe à modalidade acima, em que o campo *profundidade* adicional é introduzido ao nó sound2D. Do mesmo modo, o campo *profundidade* adicional poderia ser inserido no nó disposto de forma hierárquica acima do nó sound2D.

De acordo com uma modalidade adicional, se executa um mapeamento das coordenadas. Um campo adicional dimensionMapping no nó sound2DDepth define uma transformação, por exemplo como um vetor de 2 fileiras x 3 colunas usado para mapear o sistema de coordenadas de contexto (ccs) 2D proveniente de uma hierarquia de transformação ancestral em relação à origem do nó.

O sistema de coordenadas de nó (ncs) será calculado como se segue:

10 $ncs = ccs \times dimensionMapping$

A localização do nó trata-se de uma posição em 3 dimensões, originária da localização do vetor de entrada em 2D e da profundidade {localização.x localização.y profundidade} com relação a ncs.

15 Exemplo: o contexto de sistema de coordenadas de nós é $\{x_i, y_i\}$. dimensionMapping é $\{1, 0, 0, 0, 0, 1\}$. Isso conduz a $ncs = \{x_i, 0, y_i\}$, o que possibilita o movimento de um objeto na dimensão - y a ser mapeado em relação ao movimento de áudio na profundidade.

20 O campo "dimensionMapping" pode ser definido como MFFloat. A mesma funcionalidade também poderia ser obtida por meio do uso do tipo de dados de campo "SFRotation", que se trata de outro tipo de dado MPEG-4.

A invenção permite o posicionamento no espaço do sinal de áudio em um domínio 3D, mesmo em caso do dispositivo de reprodução estar restrito aos gráficos 2D.

REIVINDICAÇÕES

1. Método para codificação de descrição de apresentação de sinais de áudio, compreendendo as etapas de:

gerar de descrição de parâmetros de uma fonte sonora, incluindo informação que permita o posicionamento no espaço em um sistema de coordenadas 2D;

ligar a descrição de parâmetros da referida fonte sonora com os sinais de áudio da referida fonte sonora;

o método sendo **CARACTERIZADO** pelo fato de compreender a etapa de:

adicionar um valor adicional 1D à descrição de parâmetro, que permite, em um contexto visual 2D, um posicionamento no espaço da referida fonte sonora em um domínio 3D.

2. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de fontes de som separadas serem codificadas como objetos de áudio separados e a disposição das fontes de som em um cenário sonoro se descreve por uma descrição de cenário, apresentando primeiros nós, correspondendo aos objetos de áudio separados, e segundos nós, descrevendo a apresentação dos objetos de áudio, e em que um campo de um segundo nó define o posicionamento no espaço de uma fonte sonora.

3. Método, de acordo com a reivindicação 1 ou 2, **CARACTERIZADO** pelo fato de o referido sistema de coordenadas 2D corresponder ao plano de tela e o referido valor 1d corresponde a uma informação de profundidade, perpendicular ao referido plano de tela.

4. Método, de acordo com a reivindicação 3, **CARACTERIZADO** pelo fato de a transformação dos referidos valores do sistema de coordenadas 2D nas referidas posições de 3 dimensões possibilitar o movimento de um objeto gráfico no plano de tela a ser mapeado em relação a um movimento de um objeto de áudio na profundidade perpendicular ao referido plano de tela.

5. Método de decodificação de uma descrição de apresentação de sinais de áudio, compreendendo a etapa de:

10 receber de sinais de áudio correspondendo a uma fonte sonora ligada com uma descrição de parâmetros da referida fonte sonora, em que a referida descrição de parâmetros inclui informação que permite o posicionamento no espaço em um sistema de coordenadas 2D;

15 o método sendo **CARACTERIZADO** pelo fato de compreender as etapas de:

separar um valor adicional em 1D a partir da referida descrição em parâmetros; e

posicionar no espaço em um contexto visual em 2D a referida fonte sonora em um domínio 3D, utilizando o referido valor adicional em 1D.

20

6. Método, de acordo com a reivindicação 5, **CARACTERIZADO** pelo fato de que objetos de áudio representando fontes de som separadas são decodificados em separado, e uma trilha sonora simples é composta a partir dos objetos de áudio decodificados utilizando-se uma descrição de cena possuindo primeiros nós, correspondendo aos objetos de áudio separados, e segundos nós, descrevendo o processamento dos

25

objetos de áudio, e em que um campo de um segundo nó define o posicionamento no espaço 3D de uma fonte sonora.

7. Método, de acordo com a reivindicação 5 ou 6, **CARACTERIZADO** pelo fato de que o referido sistema de coordenadas 2D corresponde ao plano de tela e o referido valor em 1D corresponde a uma informação de profundidade perpendicular ao referido plano de tela.

8. Método, de acordo com a reivindicação 7, **CARACTERIZADO** pelo fato de que uma transformação dos referidos valores de sistema de coordenadas 2D nas referidas posições em 3 dimensões possibilita o movimento de um objeto gráfico no plano da tela a ser mapeado em relação a um movimento de um objeto de áudio na profundidade perpendicular ao referido plano de tela.

15 9. Dispositivo, **CARACTERIZADO** pelo fato de realizar o método do tipo definido em qualquer uma das reivindicações anteriores.