(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 411 498 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
21.04.2004 Bulletin 2004/17

(51) Int Cl.[7]: **G10L 19/00**

(21) Application number: **02022866.4**

(22) Date of filing: **14.10.2002**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: **Deutsche Thomson-Brandt GmbH**
78048 Villingen-Schwenningen (DE)

(72) Inventor: **Spille, Jens**
30966 Hemmingen (DE)

(74) Representative: **Rittner, Karsten, Dr.**
Deutsche Thomson-Brandt GmbH,
Karl-Wiechert-Allee 74
30625 Hannover (DE)

(54) Method and apparatus for describing sound sources

(57) The MPEG-4 Audio standard as defined in ISO/IEC 14496-1 and - 3 facilitates a wide variety of applications by supporting the representation of audio objects. For the combination of the audio objects additional information - the so-called scene description - determines the placement in space and time and is transmitted together with the coded audio objects.

For playback the audio objects are decoded separately and composed using the scene description in order to prepare a single soundtrack, which is then played to the listener. A scene description is structured hierarchically and can be represented as a graph, wherein nodes of the graph form the separate objects. The appearance and behaviour of the separate objects can be controlled using parameters within the scene description nodes. For describing the wideness of a sound source an audio diffuseness node is defined within the scene description.

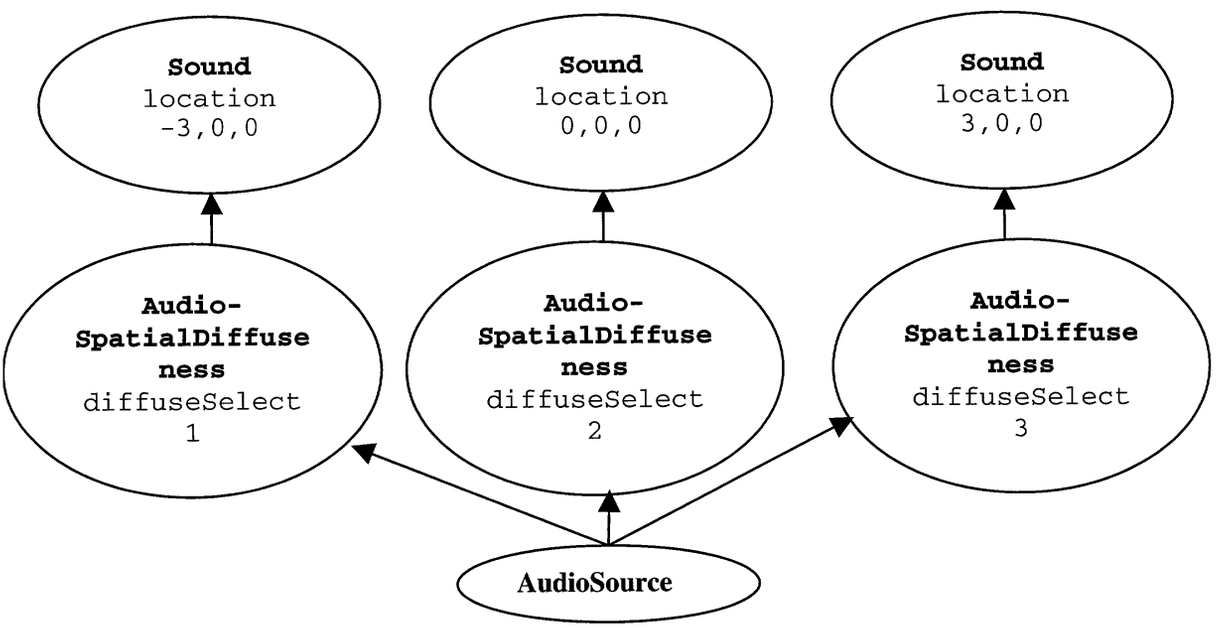

Figure 2

EP 1 411 498 A1

Description

[0001] The invention relates to a method and to an apparatus for describing sound sources, especially for sound sources encoded as audio objects according to the MPEG-4 Audio standard.

Background

[0002] The MPEG-4 Audio standard as defined in ISO/IEC 14496-3 and 14496-1 facilitates a wide variety of applications by supporting the representation of audio objects. For the combination of the audio objects additional information - the so-called scene description - determines the placement in space and time and is transmitted together with the coded audio objects.
For playback the audio objects are decoded separately and composed using the scene description in order to prepare a single soundtrack, which is then played to the listener.
[0003] For efficiency, the MPEG-4 Systems standard ISO/IEC 14496-1defines a way to encode the scene description in a binary representation, the so-called Binary Format for Scene Description (BIFS). Correspondingly, audio scenes are described using so-called AudioBIFS.
[0004] A scene description is structured hierarchically and can be represented as a graph, wherein leaf-nodes of the graph form the separate objects and the other nodes describes the processing, e.g. positioning, scaling, effects etc.. The appearance and behavior of the separate objects can be controlled using parameters within the scene description nodes.

Invention

[0005] The invention is based on the recognition of the following fact. Currently the MPEG-4 Audio standard cannot describe sound sources that have a certain dimension, like a choir, orchestra, sea or rain but only a point source, e.g. a flying insect, or a single instrument. According to listening tests wideness of sound sources is clearly audible, whereby more complicate descriptions like the shape of the audio object is not necessary.
[0006] Therefore, a problem to be solved by the invention is to allow the description of the wideness of sound sources that have a certain dimension in a simple and backwards compatible way.
[0007] This problem is solved by the method disclosed in claim 1 and the corresponding apparatus in claim 4.
[0008] In principle, the inventive method allows to describe sound sources, which are encoded as separate audio objects. The arrangement of the sound sources in a sound scene is described by a scene description. For playback the audio objects are decoded separately and a single soundtrack is composed from the decoded audio objects using said scene description. For describing the wideness of a sound source an audio spatial diffuseness node is defined within the scene description.
[0009] Advantageous additional embodiments of the invention are disclosed in the respective dependent claims.

Drawings

[0010] Exemplary embodiments of the invention are described with reference to the accompanying drawings, which show in

Fig. 1 the illustration of the functionality of the AudioSpatialDiffuseness mode;

Fig. 2 an Audio Scene for a Line Sound Source.

Exemplary embodiments

[0011] Figure 1 shows an illustration of the functionality of the inventive AudioSpatialDiffuseness node.
[0012] This AudioSpatialDiffuseness node will have a children field as input and will produce the same number of channels (numChan) as output. Branches that are connected to an upper level branch are called children in MPEG-4 terms. It can be inserted in each branch of the audio subtree, without changing any other node.
[0013] A diffuseSelection field will allow the scene author to control the diffuseness algorithms, so that each AudioSpatialDiffuseness node will produce a different output. In practice a diffuseness node will virtual produce N different signals, but only one real signal is passed through to the output of the node, signaled by the diffuseSelect field. Other fields like a decorrelation strength (decorrStrength) etc. could be added to the node, if required.

```
AudioSpatialDiffuseness  {
    eventin     MFNode  addChildren
    eventin     MFNode  removeChildren
    exposedField   MFNode  children              [ ]
    exposedField   SFInt32 diffuseSelect     1
    exposedField   SFInt32 decorreStrength   1
    field       SFInt32 numChan           1
    field       MFInt32 phaseGroup            [ ]
}
```

Table 1: Semantics of the proposed AudioSpatialDiffuseness Node

[0014] In the case of numChan greater than one each channel should be diffused separately.

[0015] Figure 2 depicts an Audio Scene for a Line Sound Source. By using this proposal the scene author has to decide how many and at which position the decorrelated multiple point sound sources will be located. The advantage is, that the content author has much more control over the shape effect. He can also use intensity and direction of each point source as well as using the AudioDelay and AudioEffects node for certain Sound nodes to manipulate the effect. It is still possible for the renderer to reduce the computational power by passing the scene tree to look for identical AudioSources.

```
# Example of a line sound source replaced by three point
sources
# using one single decoder output.

Group {
    children [
        DEF POS1 Sound {
            intensity 0.9
```

```
            location 0 0 0
            spatialize TRUE
            source AudioSpatialDiffuseness  {
                numChan 1
                diffuseSelect  1
                children [
                    DEF BEACH AudioSource {
                        numChan 1
                        url 100
                    }
                ]
            }

        DEF POS2 Sound {
            intensity 0.8
            location -3 0 0
            spatialize TRUE
            source AudioSpatialDiffuseness  {
                numChan 1
                diffuseSelect  2
                children [ USE BEACH]
            }

        DEF POS3 Sound {
            intensity 0.8
            location  3 0 0
            spatialize TRUE
            source AudioSpatialDiffuseness  {
                numChan 1
                diffuseSelect  3
                children [ USE BEACH]
            }
    ]
}
```

Table 2: Example of a Line Sound Source replaced by three Point Sources using one single Audio-Source.

Claims

1. Method for describing sound sources, which are encoded as separate audio objects, wherein the arrangement of the sound sources in a sound scene is described by a scene description, and wherein for playback the audio objects are decoded separately and a single soundtrack is composed from the decoded audio objects using said scene description, **characterized by** an audio diffuseness node which is defined within the scene description for describing the wideness of a sound source.

2. Method according to claim 1, wherein a diffuse selection field will allow the scene author to control the diffuseness algorithms.

3. Method according to claim 1 or 2, wherein a decorrelation strength field will allow author to control the strenght of the decorrelation.

4. Apparatus for performing a method according to any of claims 1 to 3.

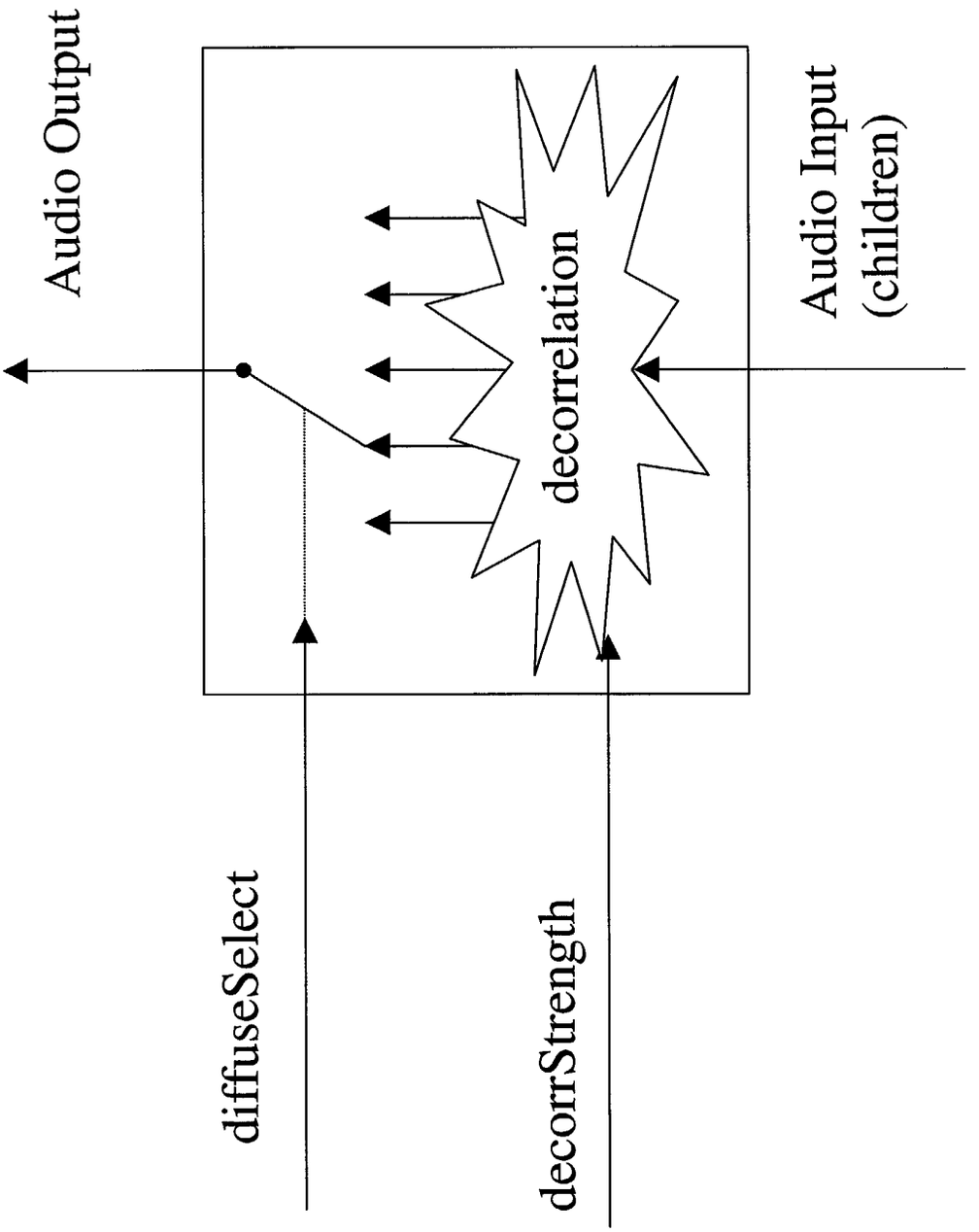
Audio Output
Audio Input
(children)
decorrelation
diffuseSelect
decorrStrength

Figure 1

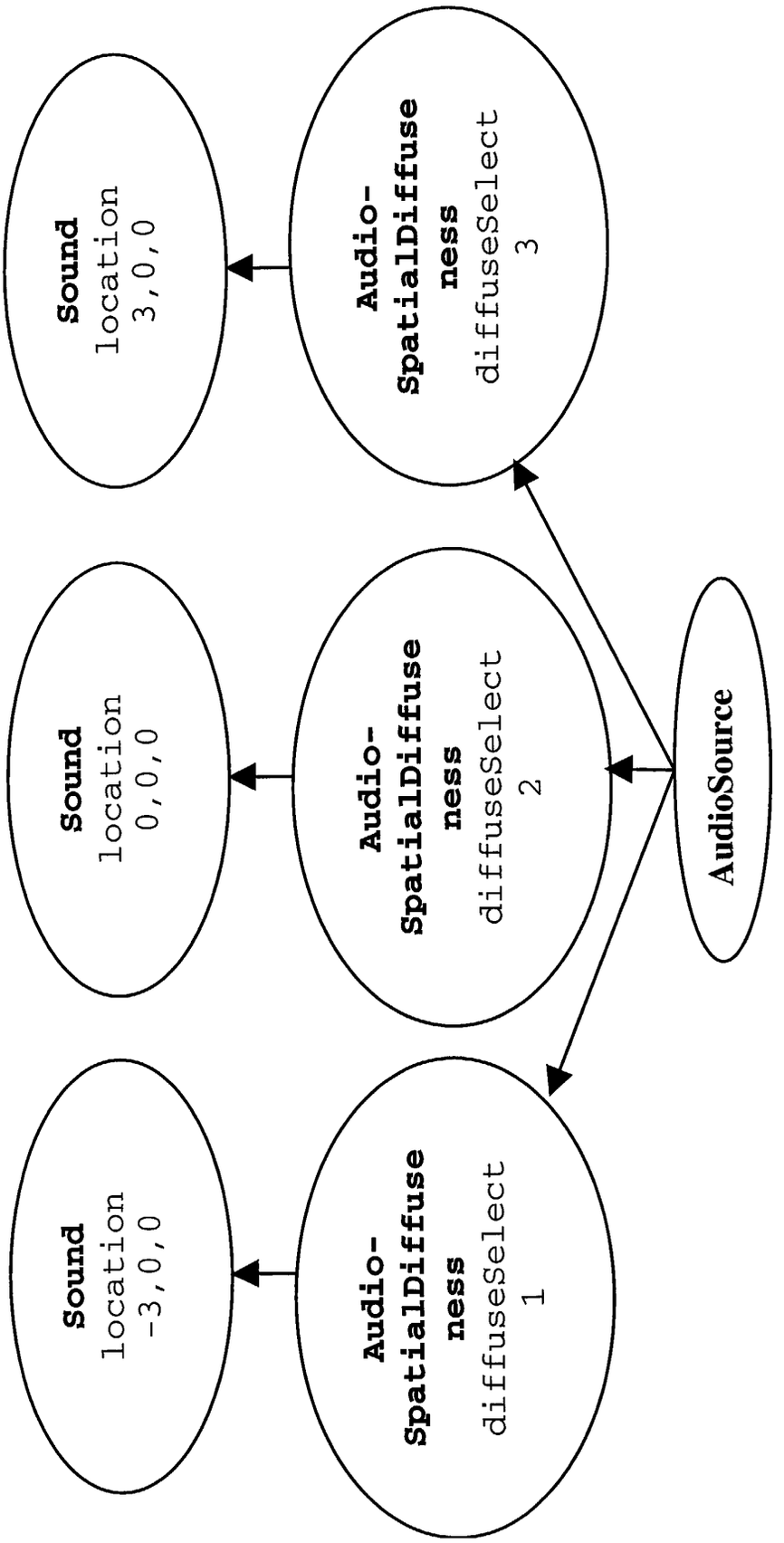
Sound
location
-3,0,0
Audio-
SpatialDiffuse
ness
diffuseSelect
1
Sound
location
0,0,0
Audio-
SpatialDiffuse
ness
diffuseSelect
2
Sound
location
3,0,0
Audio-
SpatialDiffuse
ness
diffuseSelect
3
AudioSource

Figure 2

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 02 2866

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Y	PURNHAGEN H.: "An overview of MPEG-4 audio version 2" AES 17TH INTERNATIONAL CONFERENCE ON HIGH QUALITY AUDIO CODING, 2 - 5 September 1999, XP002239258 Italy * abstract * * section "2.5. Environmental Spatialisation" * ---	1-4	G10L19/00
Y	CONVENOR: "Coding of moving pictures and audio, ISO/IEC JTC1/SC29/WG11/N4907" ORGANISATION INTERNATIONALE DE NORMALISATION, July 2002 (2002-07), XP002239259 Klagenfurt, DE * N5055 AGH on MPEG-4 Audio Extension and Exploration Core Experiments * ---	1-4	
A	BAUER S ET AL: "DER MULTIMEDIA-STANDARD MPEG-4 UND SEINE ANWENDUNGEN IM MINT-PROJEKT" FERNMELDE-INGENIEUR, BAD WINSHEIM, DE, vol. 52, no. 11/12, November 1998 (1998-11), pages 43-64, XP000997456 ISSN: 0015-010X * section "4.3 Die Komposition der Audioobjekte zu einem Klangbild * abstract * -----	1-4	TECHNICAL FIELDS SEARCHED (Int.Cl.7) G10L

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	24 April 2003	Quélavoine, R

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)