



(51) International Patent Classification:

G10L 19/008 (2013.01) H04S 3/00 (2006.01)
G10L 19/00 (2013.01) H04S 5/02 (2006.01)

(21) International Application Number:

PCT/JP2013/007390

(22) International Filing Date:

16 December 2013 (16.12.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

2013-010544 23 January 2013 (23.01.2013) JP

(71) Applicants: NIPPON HOSO KYOKAI [JP/JP]; 2-1, Jin-nan 2-chome, Shibuya-ku, Tokyo, 1508001 (JP). ELEC-TRONICS AND TELECOMMUNICATIONS RE-SEARCH INSTITUTE [KR/KR]; 218, Gajeongno, Yuseong-gu, Daejeon, 305700 (KR).

(72) Inventors: WATANABE, Kaoru; c/o NHK Science & Technology Research Laboratories, 10-11, Kinuta 1-chome, Setagaya-ku, Tokyo, 1578510 (JP). OODE, Satoshi; c/o NHK Science & Technology Research Laboratories, 10-11, Kinuta 1-chome, Setagaya-ku, Tokyo, 1578510 (JP). SAWAYA, Ikuko; c/o NHK Science & Technology Research Laboratories, 10-11, Kinuta 1-chome, Setagaya-ku, Tokyo, 1578510 (JP). YOO, Jae-hy-oun; 218, Gajeongno, Yuseong-gu, Daejeon, 305700 (KR). LEE, Taejin; 218, Gajeongno, Yuseong-gu, Daejeon,

305700 (KR). KANG, Kyeongok; 218, Gajeongno, Yuseong-gu, Daejeon, 305700 (KR).

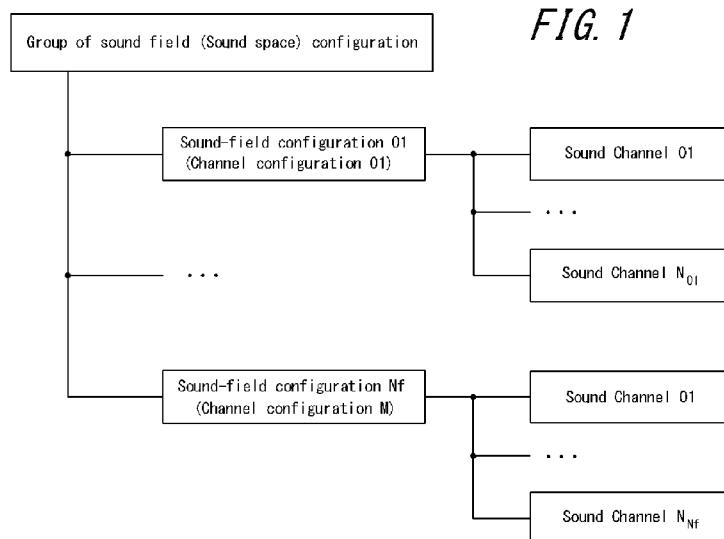
(74) Agent: SUGIMURA, Kenji; 36F, Kasumigaseki Common Gate West, 3-2-1, Kasumigaseki, Chiyoda-ku, Tokyo, 1000013 (JP).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: SOUND SIGNAL DESCRIPTION METHOD, SOUND SIGNAL PRODUCTION EQUIPMENT, AND SOUND SIGNAL REPRODUCTION EQUIPMENT



(57) Abstract: Provided is a sound signal description method corresponding to a format of "sound signals to compose a multi-layered sound field", as well as a sound signal production equipment and a sound signal reception equipment which correspond to the sound signal description method. The sound signal description method for describing the multi-layered sound field includes the number of sound field layers of the multi-layered sound field, a type of each sound field layer of the multi-layered sound field, and language information.

WO 2014/115222 A1



Published:

— *with international search report (Art. 21(3))*

Description

Title of Invention: SOUND SIGNAL DESCRIPTION METHOD, SOUND SIGNAL PRODUCTION EQUIPMENT, AND SOUND SIGNAL REPRODUCTION EQUIPMENT

Technical Field

[0001] The present invention relates to a sound signal description method, a sound signal production equipment, and a sound signal reproduction equipment, all of which are capable of representing information of sound signals with use of metadata for sound reproduction through multichannel speakers.

Background Art

[0002] Various sound systems, such as a 2 channel sound system, a 5.1 channel sound system, and "3-dimensional multichannel stereophonic sound systems" beyond the 5.1 channel sound system, are used for program production. Describing the various sound systems using a common description format provides flexibility to the sound systems, which allows the systems to be applied to next-generation sound systems across various sound application scenarios. ITU-R, which is an international standardization body associated with broadcasting including sound, has defined requirements for an advanced multichannel sound system as ITU-R Recommendation. (Refer to Non Patent Literature 1.)

Citation List

Non Patent Literature

[0003] [NPL 1] "Performance requirements for an advanced multichannel stereophonic sound system for use with or without accompanying picture", Recommendation ITU-R BS.1909.

Summary of Invention

Technical Problem

[0004] As the common description format for describing the various sound systems, an advanced study has been conducted on "sound signals to compose a single-layered sound field." However, in some cases of sound program production, the format of "sound signals to compose a multi-layered sound field" can be used so as to facilitate rendering, conversion, and switching of received sound signals according to a receiver's environment or demand of program exchange or a home reproduction. For example, the receiver of program exchange or the home sometimes does not employ the same size image display as in the program production, and according to such a video reproduction environment of the receiver, the sound signal needs to be

converted. Furthermore, it is sometimes required a language switching for program reproduction and, a reproduction position relocation of a narration signal according to needs of the receiver. Conventionally, however, the study has not been conducted on the description method for the "sound signals to compose a multi-layered sound field."

[0005] The present invention has been conceived in view of the above problems and aims to provide a sound signal description method corresponding to the format of the "sound signals to compose a multi-layered sound field", as well as a sound signal production equipment and a sound signal reproduction equipment which correspond to the sound signal description method.

Solution to Problem

[0006] In order to solve the aforementioned problems, one aspect of the present invention provides a sound signal description method for describing a multi-layered sound field, comprising: the number of sound field layers of the multi-layered sound field; a type of each sound field layer of the multi-layered sound field; and language information.

[0007] It is preferable that the type of each sound field layer of the multi-layered sound field indicates the sound elements of the program, such as one of international sound, which consists of all the sound program elements except for the commentary/dialogue elements, and one of commentary/dialogue sound with particular language.

[0008] Furthermore, another aspect of the present invention provides a sound signal description method for describing a multi-layered sound field, comprising: the number of sound field layers of the multi-layered sound field; and a video link identifier indicating, for each sound field layer of the multi-layered sound field, whether the sound field layer is linked to video.

[0009] Moreover, yet another aspect of the present invention provides a sound signal production equipment that produces a sound signal according to a sound signal description method for describing a multi-layered sound field, comprising: a metadata addition unit that produces metadata including the number of sound field layers of the multi-layered sound field, a type of each sound field layer of the multi-layered sound field, and language information; a coding unit that produces the sound signal according to the sound signal description method based on an input sound signal and the metadata; and a multiplexer that multiplexes the produced sound signal into a bit stream.

[0010] Moreover, yet another aspect of the present invention provides a sound signal reproduction equipment that reproduces a sound signal according to a sound signal description method for describing a multi-layered sound field, comprising: an environment information input unit that inputs reproduction environment and user demand information; and a rendering reproduction unit that converts the sound signal

according to the number of sound field layers of the multi-layered sound field, a type of each sound field layer of the multi-layered sound field, and language information included in the sound signal and according to the reproduction environment and user demand information, and reproduces the converted sound signal.

[0011] The type of each sound field layer of the multi-layered sound field indicates which one of international sound and a particular language the sound field layer comprises, the international sound being used irrespective of language, and the particular language being switched by the environment information input unit. The rendering reproduction unit preferably adds the sound signal of the particular language to the international sound and reproduces added sound.

[0012] Moreover, yet another aspect of the present invention provides a sound signal production equipment that produces a sound signal according to a sound signal description method for describing a multi-layered sound field, comprising: a metadata addition unit that produces metadata including the number of sound field layers of the multi-layered sound field and a video link identifier indicating, for each sound field layer of the multi-layered sound field, whether the sound field layer is linked to video; a coding unit that produces the sound signal according to the sound signal description method based on an input sound signal and the metadata; and a multiplexer that multiplexes the produced sound signal into a bit stream.

[0013] Moreover, yet another aspect of the present invention provides a sound signal reproduction equipment that reproduces a sound signal according to a sound signal description method for describing a multi-layered sound field, comprising: an environment information input unit that inputs reproduction environment information into the reproduction equipment; and a rendering reproduction unit that converts the sound signal according to the number of sound field layers of the multi-layered sound field and a video link identifier included in the sound signal and according to the reproduction environment information. The video link identifier indicating, for each sound field layer of the multi-layered sound field, whether the sound field layer is linked to video.

[0014] When the video link identifier indicates that the sound field layer is linked to video, the rendering reproduction unit preferably renders the sound signal of the sound field layer based on video display information input by the environment information input unit.

Advantageous Effects of Invention

[0015] The sound signal description method, the sound signal production equipment, and the sound signal reproduction equipment according to the present invention make it possible to describe the "sound signals to compose a multi-layered sound field" and to

produce and reproduce a sound program using the sound signals.

Brief Description of Drawings

[0016] [fig.1]FIG. 1 shows an exemplary structure of an "Extended sound field descriptor" according to one embodiment of the present invention.

[fig.2]FIG. 2 shows a block diagram of a sound signal production equipment according to one embodiment of the present invention.

[fig.3]FIG. 3 shows a block diagram of a sound signal reproduction equipment according to one embodiment of the present invention.

[fig.4]FIG. 4 is a conceptual diagram of a multi-layered sound field in connection with narration language switching.

[fig.5]FIG. 5 shows a difference in display size between a program production environment and a reproduction environment.

[fig.6]FIG. 6 is a conceptual diagram of the multi-layered sound field associated with linked/unlinked video and sound.

[fig.7]FIG. 7 shows an exemplary structure of a "Basic sound field descriptor".

Description of Embodiments

[0017] The following describes embodiments of the present invention in detail with reference to the drawings.

[0018] The present invention extends a description method (referred to below as a "Basic sound field descriptor") for describing "sound signals to compose a single-layered sound field" to the description method (referred to below as an "Extended sound field descriptor") for describing a "sound signals to compose a multi-layered sound field." Regarding the Basic sound field descriptor, the present applicants filed a Korean Patent Application (10-2012-0112984), and the Basic sound field descriptor is reviewed below for understanding of the present invention.

[0019] In order to describe multichannel sound signals to compose a single-layered sound field, it is necessary to describe which channel corresponds to the reproduction position. The described information is called descriptor, which is described as metadata in a header of a corresponding multichannel sound signal or in the headers on each sound channel constituting the multichannel.

[0020] Table 1 illustrates terms and definitions of the Basic sound field descriptor. The Basic sound field descriptor is employed for production and exchange of complete mix programs (i.e. programs including all sound required for reproduction) with multichannel sound, for example.

[Table 1]

Terms

Sound Channel	Distinct collection of sequenced sound samples that are intended for delivery to a single loudspeaker or other reproduction device. Composed of individual sound channel positions (directions) to be reproduced. Includes Type of Sound Channel Component Object (reproduction frequency level characteristics and spatial directivity characteristics). Includes an object-based signal.
Type of Sound channel component Object	Type of individual sound channel signal components (Nominal frequency-level characteristics and spatial directivity characteristics).
Sound-field configuration	Defined arrangement or configuration of loudspeakers that conveys the intended Sound-field. (A group of sound channels that are intended to be reproduced simultaneously through a defined Sound-field configuration).
Sound-field	The acoustical space within which the intended sound image is created, which are created by simultaneously reproducing sound channels described by the Sound field configuration.
Sound Essence	The sound resources that make up a sound program of television and sound-only program

[0021] The Sound Essence descriptor includes a descriptor of a program, a descriptor (name) of the Sound-field, and other relevant descriptors.

[0022] As shown in FIG. 7, the Sound-field is described by the Sound-field configuration with a hierarchical structure.

[0023] The Sound Channel descriptor includes the Channel label descriptor and/or Channel Position descriptor.

[0024] The following describes the descriptors in the Basic sound field descriptor. Note that some of the descriptors overlap with each other in anticipation of different program exchange scenarios. However, a program producer or the like is able to appropriately choose necessary descriptors for each program exchange scenario.

[0025] The Basic sound field descriptor includes (A) Sound Essence descriptors, (B) Sound-field configuration descriptors, and (C) Sound Channel descriptors.

[0026] Table 2 shows (A) Sound Essence descriptors in the Basic sound field descriptor.

[Table 2]

Name of Descriptor	Subject of Description	Example(s)
Program Name	- Program title	Program Title
Type of Sound essence (Sound-field)	- Name of Type and Content of Sound essence	Complete mix
Name of sound-field configuration	- Name of defined multichannel sound arrangement	22.2ch, 10.2ch, etc
Loudness value	- Loudness value	

[0027] Table 3 shows (B) Sound-field configuration descriptors in the Basic sound field descriptor.

[Table 3]

(B) Sound-field configuration descriptors – multichannel arrangement data

Name of Descriptor	Subject of Description	Example(s)
Name of Sound-field configuration	- Name of defined multichannel sound arrangement	22.2ch, 10.2ch, etc
The number of channels	- The total number of channel	24 channels, 12 channels
Multichannel sound arrangement description	- Numbers of horizontal and/or vertical channels	Middle:10,front:5,side:2,back:3,top:9,front:3,side:3,back:3,bottom:3,front:3,side:0,back:0,LFE:2
List of channel allocation	- Mapping of channel allocation	1:Mid_L, 2:Mid_R, 3:Mid_C, 4:LFE, 5:Mid_LS, 6:Mid_RS
Down-mixing coefficient	- Coefficients in order to down mix to conventional Sound-field (5.1ch, 2ch or 1ch)	

[0028] Table 4 shows (C) Sound Channel descriptors in the Basic sound field descriptor.

[Table 4]

(C) Sound Channel descriptors

Name of Descriptor	Subject of Description	Example(s)
Indicator of Sound Channel descriptor	- Indicator of Channel label data and Channel position data	11:Channel label data [On]/ Channel position data [On]

[0029] Table 5 shows C.1 Channel label descriptors, which are descriptors of the Channel label data included in the Sound Channel descriptors.

[Table 5]

C.1 Channel label descriptors

Name of Descriptor	Subject of Description	Example(s)
Allocation number	- Allocation number	1:first channel, 2:second channel, etc
Channel label (A label to indicate the intended channel for sound reproduction)	- Horizontal Channel label	C:Center of screen, L:Left side of screen, Lc:Inner side on the left of the screen, Lw:Outer side on the left of screen
	- Vertical Channel label	Mid:Middle layer, Tp:Top layer (above the listener's ear height), Bt:Bottom layer (under the listener's ear height)
	- Distance Channel label	Near, Far
	- Object Channel label	Vocal, Piano, Drum, etc
Type(Characteristics) of channel component object	- Nominal frequency Range	Full:general channel, LFE:Low frequency effect channel (Include channel label or other?)
	- Type of channel component directivity	/Direct/Diffuse/Surround (Include channel label or other?)
	- Moving Information	Information for moving objects: (Time, position) information

[0030] Table 6 shows C.2 Channel position descriptors, which are descriptors of the Channel position data included in the Sound Channel descriptors.

[Table 6]

C.2 Channel position descriptors

Name of Descriptor	Subject of Description	Example(s)
Allocation number	- Allocation number	1: first channel
Spatial position data	- Azimuth angle	000:center of screen, 060:60-degrees
	- Elevation angle	000:position of listener's ear height, 060:60-degrees
Distance position data	- distance	3:3meter
Tolerance of Spatial position	- horizontal tolerance	10:±10 degrees, 15:±15 degrees
	- vertical tolerance	10:±10 degrees, 15:±15 degrees
	- Moving Information of time	Information for moving objects: especially Time information
Tolerance of Distance position	- distance	3:3 meter
	- Moving Information of position	Information for moving objects: especially Position information
Type (Characteristics) of channel component object	- Nominal frequency Range	Full: general channel, LFE: Low frequency effect channel
	- Type of channel component directivity	/Direct/Diffuse/Surround

[0031] The present invention extends the Basic sound field descriptor, which is the description method for the "sound signals to compose a single-layered sound field" as mentioned above, to the Extended sound field descriptor, which is the description method for the "sound signals to compose a multi-layered sound field."

[0032] Table 7 illustrates terms and definitions of the Extended sound field descriptor.

[Table 7]

Terms

Sound Essence	The sound resources that make up a sound program of television and sound-only program.
Group of sound field configurations (Sound space configurations)	A group of one or more Sound field configurations which are meant to be transmitted simultaneously. A group of Sound-field configurations which are intended to be (possibly) reproduced simultaneously through a defined Layered-Sound-field configuration. Example: Sound field of dialogue + Sound field of SE
Sound-field	The acoustical space within which the intended sound image is created, which is created by simultaneously reproducing sound channels described by the Group of sound field configurations.
Sound-field configuration	Defined arrangement or configuration of loudspeakers that conveys the intended Sound-field. (A group of sound channels that are intended to be reproduced simultaneously through a defined Sound-field Configuration).
Sound field of Spatial anchor (SE)	Sound field consisting of Spatial anchor (SE) element / Indicate of Spatial anchor (SE) Sound field.
Sound field of Dialogue	Sound field consisting of Dialogue element / Indicate of Dialogue Sound field.
Sound field of Video linked objects	Sound field of television program and the Sound field linked to Video signals.
Sound Channel	Distinct collection of sequenced sound samples that are intended for delivery to a single loudspeaker or other reproduction equipments. Composed of individual sound channel positions (directions) to be reproduced. Includes Type of Sound Channel Component Object (reproduction frequency level characteristics and spatial directivity characteristics). Includes an object-based signal.

[0033] The Sound Essence descriptor includes the descriptor of the program, the descriptor (name) of the Sound-field, and the other relevant descriptors.

[0034] As shown in FIG. 1, the Sound-field in the Extended sound field descriptor is described by multiple Sound-field configurations (Group of sound-field configurations) (Sound space configurations) each having the hierarchical structure.

[0035] The Sound Channel descriptor includes the Channel label descriptor and/or the Channel Position descriptor.

[0036] Table 8 shows (A) Sound Essence descriptors in the Extended sound field descriptor.
[Table 8]

(A) Sound Essence descriptors (incl. Sound field)

Name of Descriptor	Subject Description of	Example(s)
Program name	– Program name	Programme Title,
The number of Sound-field layers	– The total number of Sound-field layers	2
List of Sound-field layers and Sound-field layer Type	– List of Sound-field layers and Sound-field layer Type	complete mix, international mix, spatial anchor, dialogue, commentary, music, sound effects, hearing impaired, visual impaired, video linked objects, [Samples] 01 spatial anchor, 02 video linked objects, 03 dialogue

[0037] Table 9 shows A.2 Sound-field descriptors in the Extended sound field descriptor.

[Table 9]

A.2 Sound-field descriptors (each layer)

Name of Descriptor	Subject of Description	Example(s)
Sequential number of Sound-field	– Sequential number	1
Type of Sound-field layer	– Name of Type and Content of Sound-field	complete mix, international mix, spatial anchor, dialogue, commentary, music, sound effects, hearing impaired, visual impaired, video linked objects
Video link indicator	– Linked / unlinked	linked
Description of video format/viewing angle	– Type of video format – video viewing angle	without video, SD, HD, UHDTV(4k), UHDTV(8k) horizontal viewing angle (degree) 100°
Name of Sound field configuration	– Name of defined multichannel sound arrangement or configuration	22.2ch, 10.2ch, etc.
Language	– Language	Korean, Japanese, Null,

[0038] Regarding (B) Sound-field configuration descriptors and (C) Sound Channel descriptors in the Extended sound field descriptor, these descriptors are the same as those of the Basic sound field descriptor, and a description thereof is omitted.

[0039] FIG. 2 shows a block diagram of a sound signal production equipment according to one embodiment of the present invention. In order to "facilitate" rendering, conversion, and switching of received sound signals according to the receiver's environment or demand of program exchange or the home reproduction, the sound signal production equipment produces a sound program according to the Extended sound field descriptor, which is the format of the "sound signals to compose a multi-layered sound field." The sound signal production equipment inserts the Extended sound field descriptor as metadata into the header of the corresponding sound format signal or into the header of each audio signal, for program exchange and transmission to the home. The sound signal production equipment includes a mixing unit 11, a metadata addition unit 12, a coding unit 13, a multiplexer 14, and a monitoring unit 15.

[0040] The mixing unit 11 mixes sound signals (Sound Sources 1-M) and outputs, to the coding unit 13, sound signals to compose the multi-layered sound field including Spatial anchor, Commentary, Dialogue, and Object signals, the sound signals being output from a "production system for sound signals to compose a multi-layered sound field."

- [0041] The metadata addition unit 12 outputs, to the coding unit 13, the metadata to be described for the Extended sound field descriptor of the multi-layered sound field including Spatial anchor, Commentary, Dialogue, and Object signals. The metadata addition unit 12 also outputs the produced metadata to the coding unit 13.
- [0042] Based on the mixed sound signals received from the mixing unit 11 and the metadata received from the metadata addition unit 12, the coding unit 13 produces the sound signals according to the Extended sound field descriptor, encodes the produced sound signals, and outputs the encoded sound signals to the multiplexer 14.
- [0043] The multiplexer 14 receives, from the coding unit 13, the sound signals according to the Extended sound field descriptor that have been encoded, and multiplexes the received sound signals into a bit stream, in order to convey a multiplexed sound signal to a sound signal reproduction equipment via broadcast or transmission. The multiplexer 14 transmits the multiplexed bit stream to remote places such as home via radio waves, IP circuits, and the like.
- [0044] The monitoring unit 15 is used for checking contents of the sound signals and the metadata.
- [0045] FIG. 3 shows a block diagram of the sound signal reproduction equipment according to one embodiment of the present invention. In accordance with an input of information about a reproduction system, such as speaker arrangement information and user demand of narration sound position to be reproduced, the sound signal reproduction equipment utilizes the metadata included in the received sound signal and reproduces the received sound signal by controlling narration sound to be adjusted to a narration language and narration reproduction position desired by a user, while maintaining high quality sound providing as much of a sense of presence as was produced. Furthermore, in a reproduction environment with a video display having a different size from a size according to production conditions, the sound signal reproduction equipment controls a sound image field position in the sound field layer of a "video/sound linked sound source", which requires a link between video and sound image positions, to be adjusted to the video display, and reproduces sound appropriately for reproduction environment with the video display, while maintaining the high quality sound providing as much of the sense of presence as was produced. The sound signal reproduction equipment includes a demultiplexer 21, a decoding unit 22, a rendering reproduction unit 23, an environment information input unit 24, and monitoring unit 25.
- [0046] The demultiplexer 21 receives, via broadcast or transmission, the sound signal according to the Extended sound field descriptor that has been multiplexed into the bit stream, and demultiplexes the received sound signal into the respective sound signals of the sound field layers and the metadata. The demultiplexer 21 also outputs the de-

multiplexed sound signals and metadata to the decoding unit 22.

- [0047] The decoding unit 22 decodes the encoded sound signals and metadata received from the demultiplexer 21 and outputs, to the rendering reproduction unit 23, signals including Spatial anchor, Commentary, Dialogue, Object signals, and metadata.
- [0048] Based on the Extended sound field descriptor, the rendering reproduction unit 23 reproduces the original sound signals as they are, or renders (e.g. down-mixes) the sound signals based on the reproduction environment (e.g. the number of channels of a speaker and a display size) before reproducing the sound signals. That is to say, the rendering reproduction unit 23 renders (e.g. switches, converts, and renders) the sound signals based on the Extended sound field descriptor in a sound reproduction environment different from the environment during program production.
- [0049] The environment information input unit 24 displays to a user the metadata information described as the Extended sound field descriptor, receives user inputs about the reproduction environment and demand information of the user, namely, language selection for the multiplexed sound, reproduction environment information (e.g. the speaker configuration and the display size), and the like, and outputs the input information to the rendering reproduction unit 23.
- [0050] The monitoring unit 25 is used for checking a result of reproduction performed by the rendering reproduction unit 23, as well as program viewing.
- [0051] The following describes specific usage embodiments of the sound signal production equipment and the sound signal reproduction equipment. For example, the sound signal production equipment and the sound signal reproduction equipment according to the present invention make it possible to easily control the narration language switching and narration reproduction position relocation in accordance with the home reproduction environment and user demand. Furthermore, in the reproduction environment with the video display having the different size than the size according to production conditions, the sound signal production equipment and the sound signal reproduction equipment according to the present invention make it possible to easily control the sound image field position in the sound field layer of the "video/sound linked sound source", which requires the video to be linked to the sound image position, to be adjusted to the video display and perform reproduction, while maintaining the high quality sound providing as much of the sense of presence as was produced.
- [0052] (Production Embodiment 1: Production of Signal Including Sound Field Layer Associated with multiple Languages)
- As an example of program production using the Extended sound field descriptor, i.e., the format of the "sound signals to compose a multi-layered sound field", suppose a case where not only the sound signals of the Japanese or Korean narrations and

dialogues but also the sound signals of various languages such as English are produced. In the above example, the sound signal production system is formed by the format of the "sound signals to compose a multi-layered sound field" including the sound field layer of the international sound (Spatial anchor) used irrespective of language, and the sound field layers (Commentary, Dialogue) of the narrations and dialogues of particular languages.

[0053] In this example, the metadata addition unit 12 adds the metadata shown in Table 10 to the header of the corresponding multichannel-sound-format signal or to the headers on each sound channel constituting the multichannel according to the Extended sound field descriptor.

[Table 10]

Name	Function
The number of layers of sound field (A: The number of Sound-field layers)	Indicates how many sound field layers are included.
Sound field layer type (A.2: Type of Sound-field)	Indicates the type of each sound field layer, such as international sound and dialogue.
Language information (A.2: Language)	Indicates the languages of dialogue and narration sound field layers.

[0054] (Reproduction Embodiment 1: Reproduction of Signal Including Sound Field Layer Associated with Multiple Languages)

The user inputs the information of the reproduction system, such as the speaker arrangement information and the user demand of narration sound position to be reproduced, and controls the sound signals (e.g. the user arbitrarily adjusts the reproduction position). For example, in the home reproduction environment the sound signals can be reproduced under control in terms of a desired narration language and narration reproduction position while the high quality sound providing as much of the sense of presence as was produced is maintained.

[0055] In order to achieve the above function, the user at a receiving side inputs, through the environment information input unit 24, the information of desired narration sound (e.g. the narration language that the user demands to reproduce and the narration reproduction position) and the information of the reproduction system (e.g. speaker arrangement information). The rendering reproduction unit 23 switches a sound signal of the "narration language" layer that has been designated from among the produced narration languages described in the metadata, adds to the switched sound signal the international sound used irrespective of language for reproduction, and reproduces the sound signal. The rendering reproduction unit 23 is also fed the desired narration reproduction position, the speaker arrangement information, and the sound signal of the

produced "narration language" layer. The rendering reproduction unit 23 also relocates the switched sound signal so that reproduction is performed from the designated narration reproduction position and renders the signal so that the sound quality providing as much of the sense of presence as was produced is achieved. Subsequently, the rendering reproduction unit 23 adds, to the rendered signal, the international sound used irrespective of language and reproduces the signal.

[0056] FIG. 4 is a conceptual diagram of the multi-layered sound field including the sound field layer of the international sound (Spatial anchor) used irrespective of language, and the sound field layers of the "narration languages" (Commentary, Dialogue).

[0057] (Production Embodiment 2: Production of Program Including Sound Field Layer Associated with Linked/Unlinked Video and Sound)

As an example of program production using the Extended sound field descriptor, i.e., the format of the "sound signals to compose a multi-layered sound field", suppose a case where the "sound requiring the link between video and sound positions" and the "sound directly irrespective of the video position" are separately produced and recorded. Sound signals include not only the "sound requiring the link between video and sound positions" (e.g. the dialogue of an actor and sound emitted from an object on the screen) but also the "sound directly irrespective of the video position" (e.g. sound effects for enhancing the sense of presence of an entire program), and the "sound requiring the link between video and sound positions" and the "sound directly irrespective of the video position" can be separately produced and recorded. In the above example, the sound signal production system is formed by the format of the "sound signals to compose a multi-layered sound field" including the sound field layer of the "sound requiring the link between video and sound positions" and the "sound directly irrespective of the video position."

[0058] In this example, the metadata addition unit 12 adds the metadata shown in Table 11 to the header of the corresponding multichannel sound format signal or to the headers on each sound channel constituting the multichannel according to the Extended sound field descriptor.

[Table 11]

Name	Function
The number of layers of sound field (A: The number of Sound-field layers)	Indicates how many sound field layers are included.
Video Link Identifier (A.2: Video link indicator)	Indicates whether or not the sound field layer is linked to video.
Video format/viewing angle (A.2: Description of video format/viewing angle)	Indicates the type of video format and an optimal viewing angle in the sound field linked to video.

[0059] (Reproduction Embodiment 2: Reproduction of Program Including Sound Field Layer Associated with Linked/Unlinked Video and Sound)

In the reproduction environment with the video display having the different size than the size according to the production conditions as shown in FIG. 5, for example, the sound signal reproduction equipment controls the sound image field position in the sound field layer of the "video/sound linked sound source", which requires the link between video and sound image positions, to be adjusted to the video display and reproduces sound, while maintaining the high quality sound providing as much of the sense of presence as was produced.

[0060] In order to achieve the above function, the user at the receiving side inputs, through the environment information input unit 24, the information of the reproduction system (e.g. speaker arrangement and video display information). When the conditions for the video display and the speaker arrangement during production are the same as the conditions for the video display and the speaker arrangement at the receiving side, the rendering reproduction unit 23 does neither convert nor render the received sound signals. In this case, the rendering reproduction unit 23 adds the "sound requiring the link between video and sound positions" and the "sound directly irrespective of the video position" and reproduces the added sound. On the other hand, when the above conditions are not the same in terms of either one of the video display and the speaker arrangement, the rendering reproduction unit 23 converts the received sound signals by either rendering or down-mixing so that the sound quality providing as much of the sense of presence as was produced is achieved, and reproduces the added sound signals. When the video display size is different, and the speaker arrangement is the same, the rendering reproduction unit 23 renders the sound signals of the layer of the "sound preferably requiring the link between video and sound positions" so that a width of the video display size equals a width of the sound image. The rendering reproduction unit 23 adds the rendered "sound preferably requiring the link between video and sound positions" and the unconverted and un-rendered "sound directly irrespective of the video position" and reproduces the added sound. Here, the rendering processing, i.e., processing for equalizing the width between the sound image of the "sound preferably requiring the link between video and sound positions" and the video display size, can be easily performed by using field position information of Azimuth angle and Elevation angle included in Spatial position data defined in Channel position data.

[0061] FIG. 6 is a conceptual diagram of the multi-layered sound field including the sound field layer of "video/sound linked sound source" (Video linked object) and the sound field layers "directly irrespective of the video position" (Spatial anchor, Dialogue).

[0062] Thus, according to the above embodiment, the Extended sound field descriptor

includes the number of sound field layers, the type of each sound field layer, and the language information. With the above structure, the sound signal description method corresponding to the format of the "sound signals to compose a multi-layered sound field" is achieved.

- [0063] Furthermore, it is preferable that the type of each sound field layer indicates which one of international sound and a particular language the sound field layer comprises, the international sound being used irrespective of language. With the above structure, in the home reproduction environment, for example, the sound signals can be re-produced under control in terms of the desired narration language and narration reproduction position while the high quality sound providing as much of the sense of presence as was produced is maintained.
- [0064] Moreover, according to the above embodiment, the Extended sound field descriptor includes the number of multiple sound field layers and a video link identifier indicating, for each sound field layer, whether the sound field layer is linked to video. With the above structure, in the reproduction environment with the video display having the different size than the size according to the production conditions, for example, the sound image field position in the sound field layer of the "video/sound linked sound source", which requires the link between video and sound image positions, can be controlled to be adjusted to the video display, and reproduction is performed, while the high quality sound providing as much of the sense of presence as was produced is maintained.
- [0065] Moreover, with the sound signal production equipment and the sound signal reproduction equipment according to the above embodiments, the sound signal described by the Extended sound field descriptor can be produced and reproduced. Note that the present invention also includes, in its scope, any equipment that transmits the sound signal described by the Extended sound field descriptor to the remote places such as home via radio waves, IP circuits, and the like, any equipment that stores and records in a recording medium the sound signal described by the Extended sound field descriptor, and a recording medium in which the sound signal described by the Extended sound field descriptor is stored and recorded.
- [0066] The sound signal production equipment according to one embodiment of the present invention produces the metadata including the number of sound field layers, the type of each sound field layer, and the language information, produces the sound signal according to the Extended sound field descriptor based on an input sound signal and the metadata, and multiplexes the sound signal into the bit stream. Furthermore, the sound signal reproduction equipment according to one embodiment of the present invention converts the sound signal according to the number of sound field layers, the type of each sound field layer, and the language information included in the sound

signal and according to the reproduction environment and demand information of the user, and reproduces the converted sound signal. The above structure makes it possible to produce and view a program using the "sound signals to compose a multi-layered sound field." In particular, the sound signal reproduction equipment adds, to the international sound, the sound signal of the particular language that has been switched by the user, and reproduces the added sound. The above structure allows the user to arbitrarily carry out an operation such as language selection with use of the received metadata, thereby making it possible to switch and relocate the appropriate narration language and narration reproduction position, while the high quality sound providing as much of the sense of presence as was produced is maintained.

[0067] Moreover, the sound signal production equipment according to one embodiment of the present invention produces the metadata including the number of layers of sound field and a video link identifier indicating, for each sound field layer, whether the sound field layer is linked to video, produces the sound signal according to the Extended sound field descriptor based on the input sound signal and the metadata, and multiplexes the sound signal into the bit stream. Moreover, the sound signal reproduction equipment according to one embodiment of the present invention converts the sound signal according to the video link identifier and according to the reproduction environment information of the user, the video link identifier indicating, for each sound field layer, whether the sound field layer is linked to video, and the sound signal reproduction equipment reproduces the converted sound signal. The above structure makes it possible to produce and view the program using the "sound signals to compose a multi-layered sound field." In particular, when the video link identifier indicates that the sound field layer is linked to video, the rendering reproduction unit renders the sound signal of the sound field layer based on information about the video display of the user, and reproduces the rendered sound signal. The above structure makes it possible to render and convert the sound image field position in the sound field layer of the "video/sound linked sound source", which requires the link between video and sound image positions, so that the sound field image position is adjusted to the video display, while the high quality sound providing as much of the sense of presence as was produced is maintained by inputting the information of the reproduction system (e.g. the video display) of the user and by using the information of the video display during production described in the metadata.

[0068] While the present invention has been described based on the figures and embodiments, it should be noted that a person skilled in the art can readily make various modifications and changes in accordance with the disclosure. As such, it should also be noted that the modifications and changes are within the scope of the present invention. For example, the function or the like included in each element, each means, and each

step is subject to rearrangement, and several means and steps can be combined into a single means or step or they can be divided.

Industrial Applicability

[0069] The present invention makes it possible to describe a "sound signals to compose a multi-layered sound field", and to produce and view/listen a program using such sound signals. As a result, interoperability between different next generation sound systems is achieved, and even in a sound reproduction environment different from the environment during program production, switching, conversion, and rendering of the sound signals is facilitated.

Reference Signs List

[0070] 11 mixing unit
12 metadata addition unit
13 coding unit
14 multiplexer
15 monitoring unit
21 demultiplexer
22 decoding unit
23 rendering reproduction unit
24 environment information input unit
25 monitoring unit

Claims

- [Claim 1] A sound signal description method for describing a multi-layered sound field, comprising:
the number of sound field layers of the multi-layered sound field;
a type of each sound field layer of the multi-layered sound field; and
language information.
- [Claim 2] The sound signal description method recited in claim 1, wherein the type of each sound field layer of the multi-layered sound field indicates which one of international sound and a particular language the sound field layer comprises, the international sound being used irrespective of language.
- [Claim 3] A sound signal description method for describing a multi-layered sound field, comprising:
the number of sound field layers of the multi-layered sound field; and
a video link identifier indicating, for each sound field layer of the multi-layered sound field, whether the sound field layer is linked to video.
- [Claim 4] A sound signal production equipment that produces a sound signal according to a sound signal description method for describing a multi-layered sound field, comprising:
a metadata addition unit that produces metadata including the number of sound field layers of the multi-layered sound field, a type of each sound field layer of the multi-layered sound field, and language information;
a coding unit that produces the sound signal according to the sound signal description method based on an input sound signal and the metadata; and
a multiplexer that multiplexes the produced sound signal into a bit stream.
- [Claim 5] A sound signal reproduction equipment that reproduces a sound signal according to a sound signal description method for describing a multi-layered sound field, comprising:
an environment information input unit that inputs reproduction environment and user demand information; and
a rendering reproduction unit that converts the sound signal according to the number of sound field layers of the multi-layered sound field, a type of each sound field layer of the multi-layered sound field, and

language information included in the sound signal and according to the reproduction environment and user demand information, and reproduces the converted sound signal.

[Claim 6] The sound signal reproduction equipment recited in claim 5, wherein the type of each sound field layer of the multi-layered sound field indicates which one of international sound and a particular language the sound field layer comprises, the international sound being used irrespective of language, and the particular language being switched by the environment information input unit, and the rendering reproduction unit adds the sound signal of the particular language to the international sound and reproduces added sound.

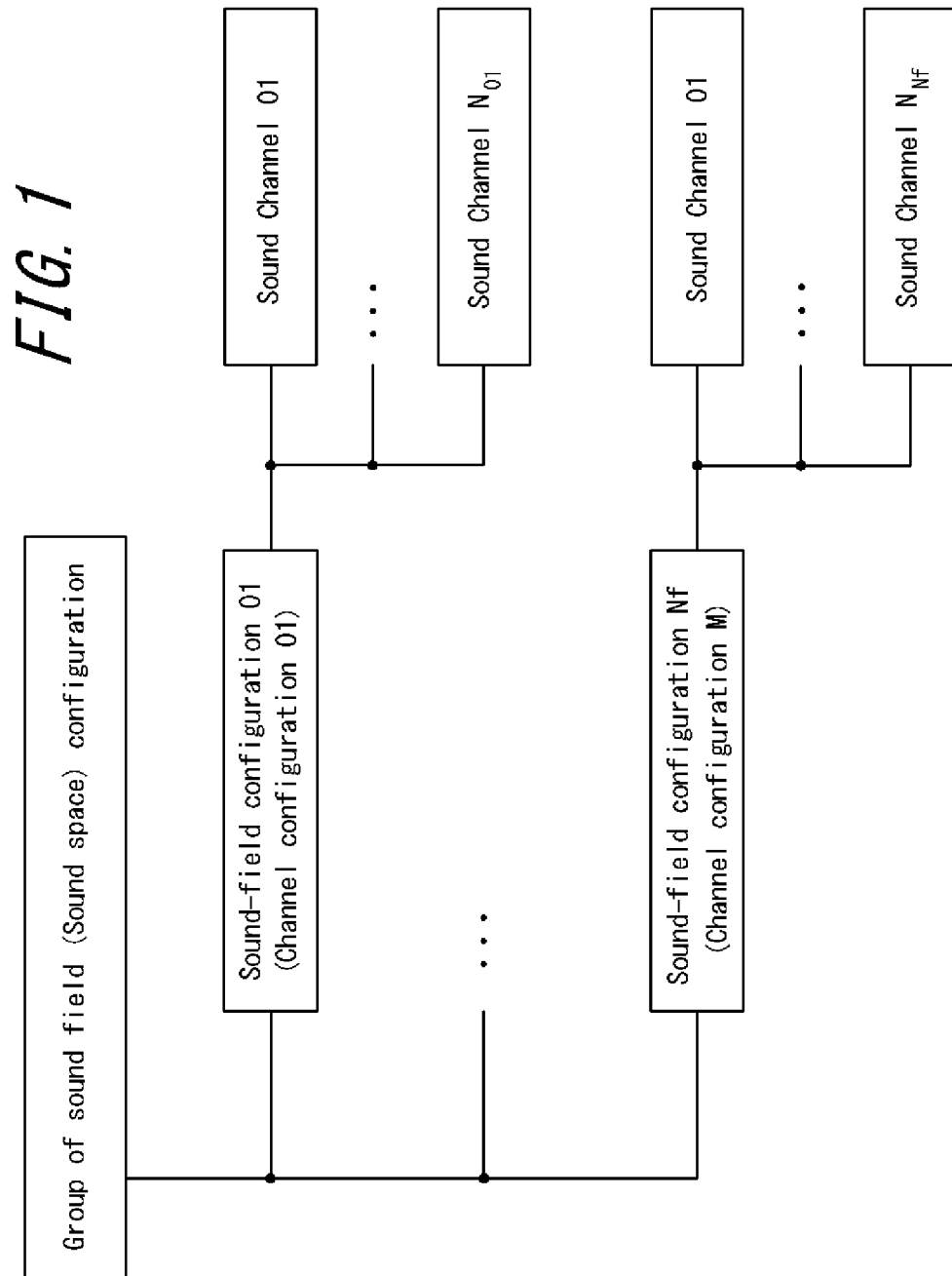
[Claim 7] A sound signal production equipment that produces a sound signal according to a sound signal description method for describing a multi-layered sound field, comprising:
a metadata addition unit that produces metadata including the number of sound field layers of the multi-layered sound field and a video link identifier indicating, for each sound field layer of the multi-layered sound field, whether the sound field layer is linked to video;
a coding unit that produces the sound signal according to the sound signal description method based on an input sound signal and the metadata; and
a multiplexer that multiplexes the produced sound signal into a bit stream.

[Claim 8] A sound signal reproduction equipment that reproduces a sound signal according to a sound signal description method for describing a multi-layered sound field, comprising:
an environment information input unit that inputs reproduction environment and user demand information; and
a rendering reproduction unit that converts the sound signal according to the number of sound field layers of the multi-layered sound field and a video link identifier included in the sound signal and according to the reproduction and user demand environment information, the video link identifier indicating, for each sound field layer of the multi-layered sound field, whether the sound field layer is linked to video.

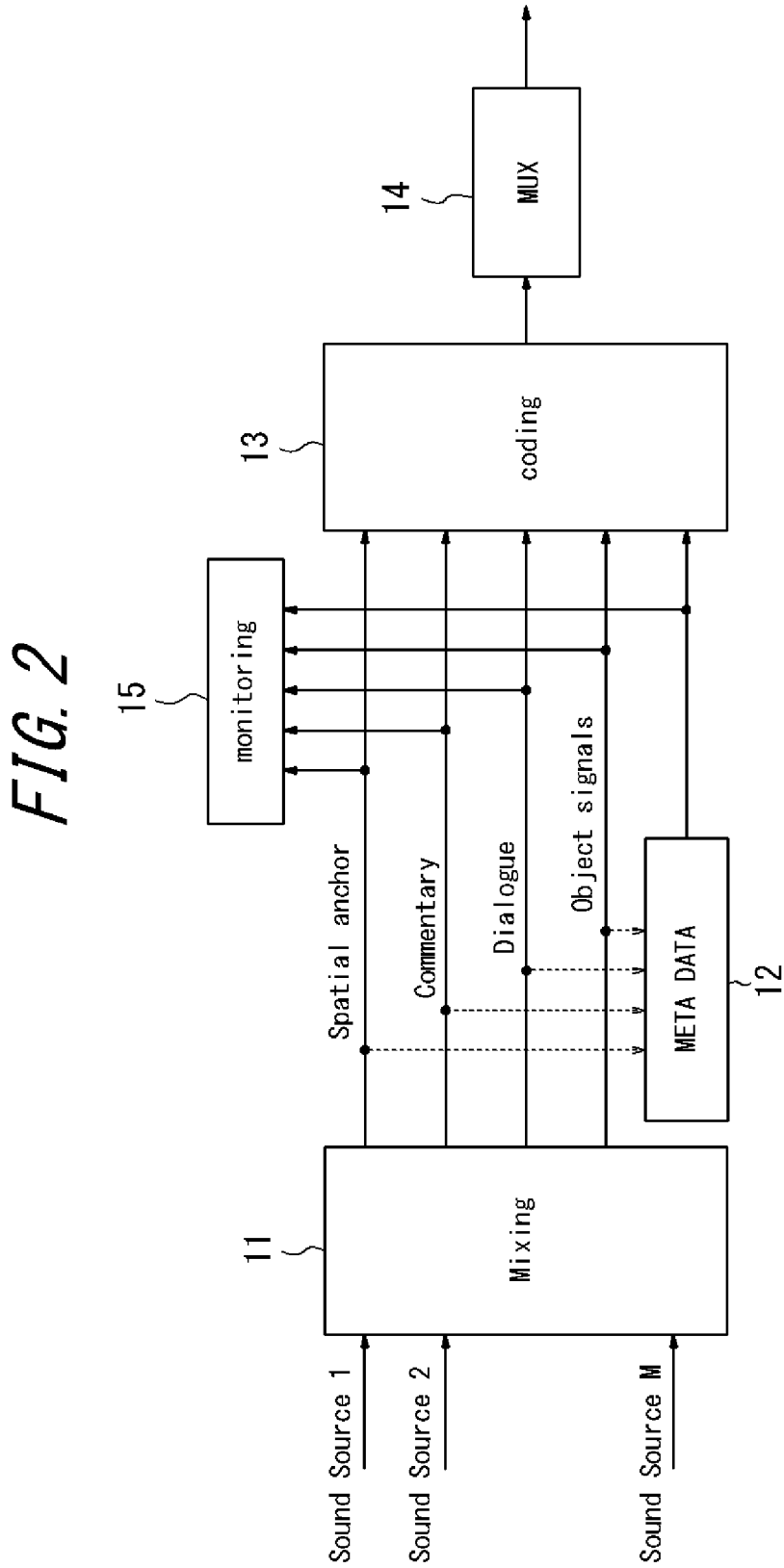
[Claim 9] The sound signal reproduction equipment recited in claim 8, wherein when the video link identifier indicates that the sound field layer is linked to video, the rendering reproduction unit renders the sound signal of the sound field layer based on video display information input

by the environment information input unit.

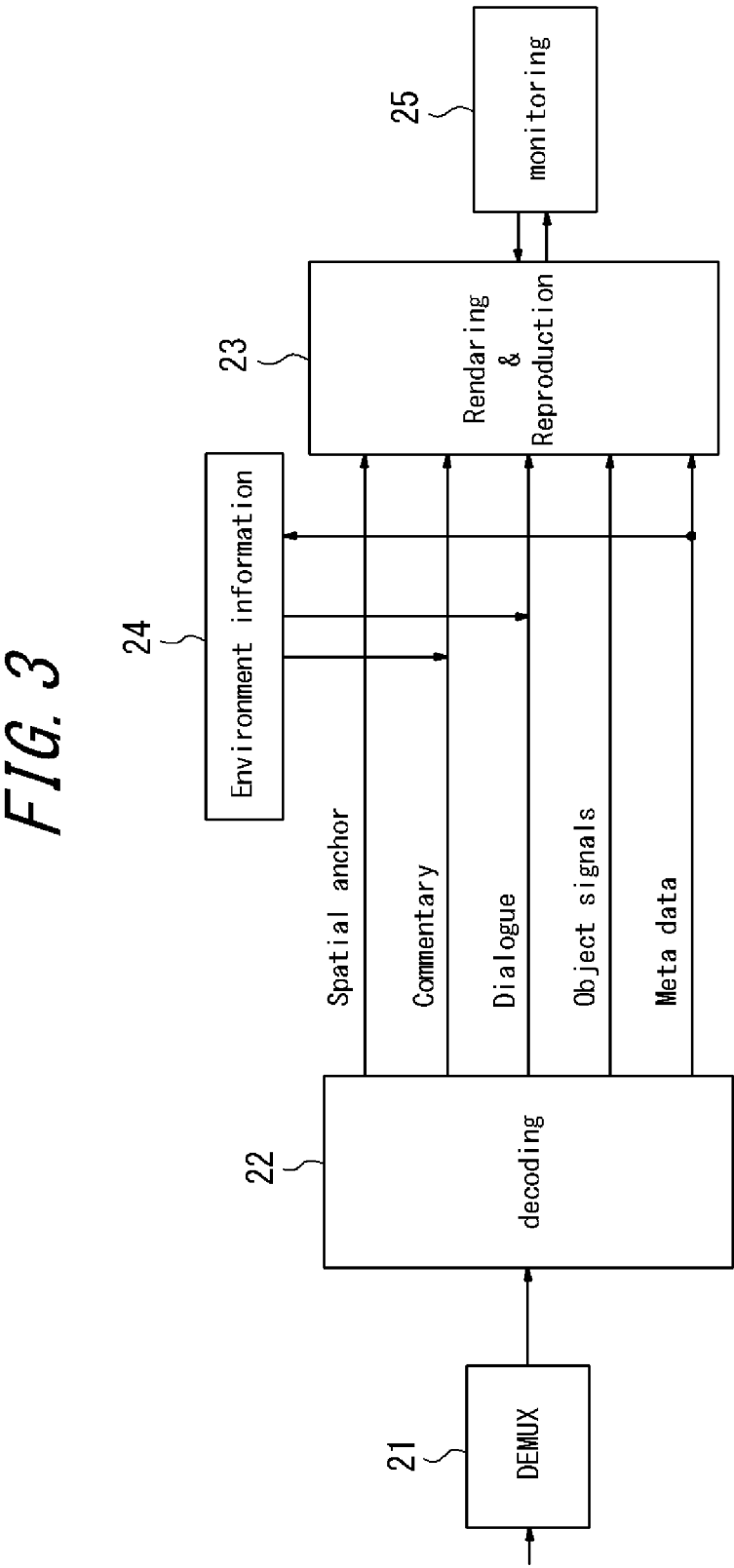
[Fig. 1]



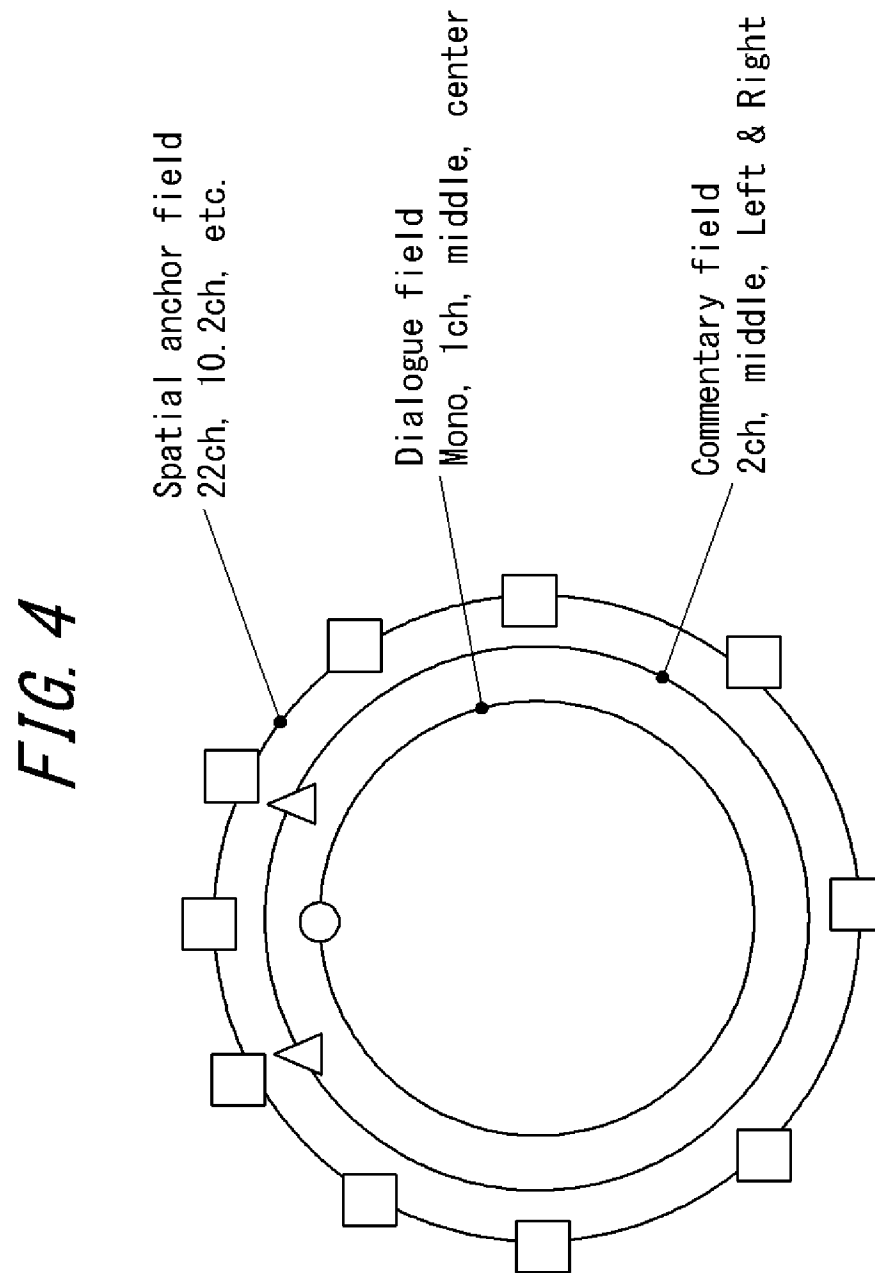
[Fig. 2]



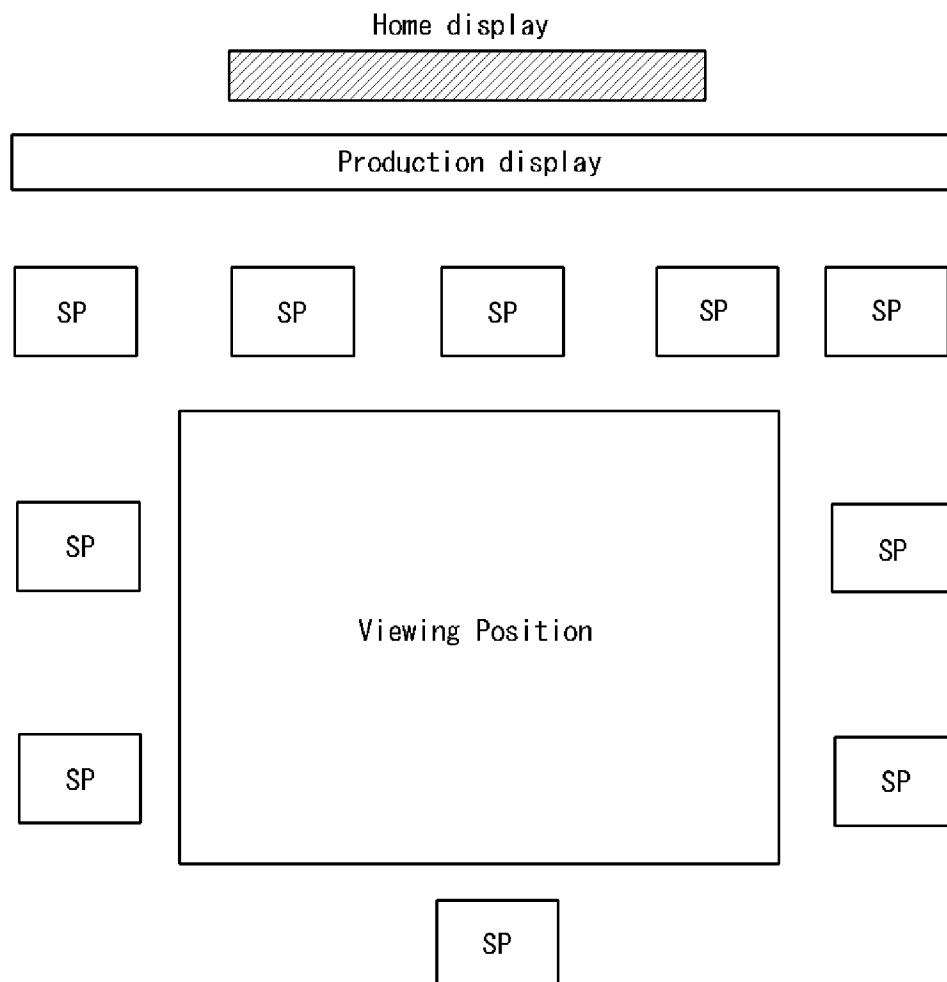
[Fig. 3]



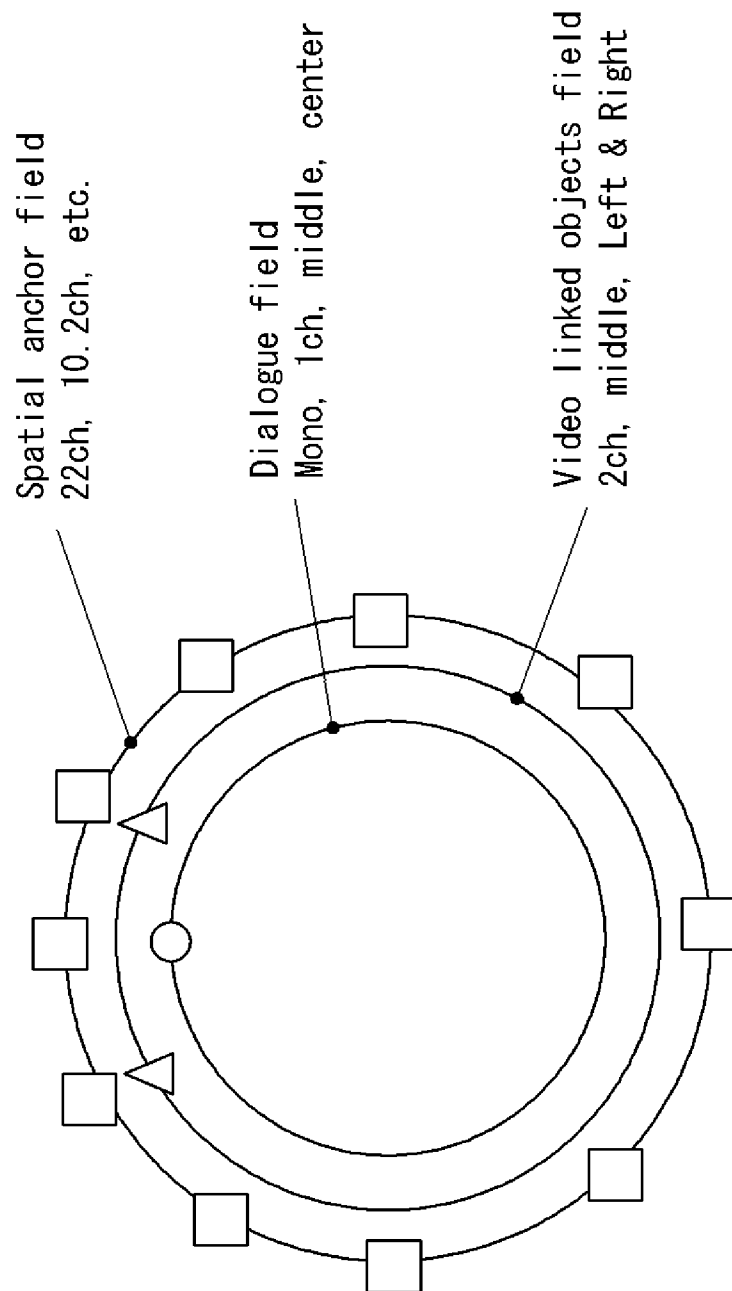
[Fig. 4]



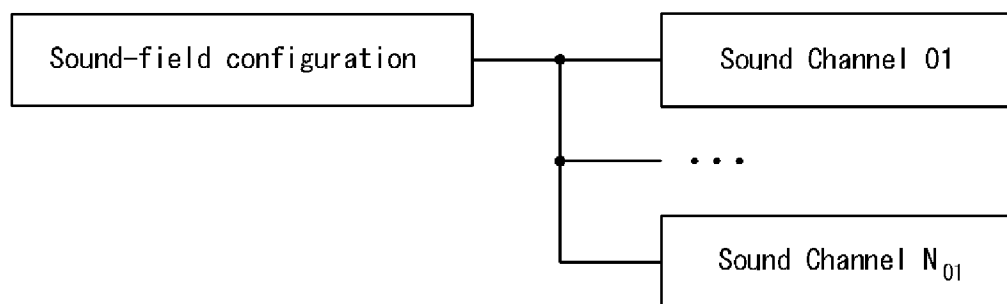
[Fig. 5]

FIG. 5

[Fig. 6]

FIG. 6

[Fig. 7]

FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/007390

A. CLASSIFICATION OF SUBJECT MATTER			
Int.Cl. G10L19/008 (2013.01) i, G10L19/00 (2013.01) i, H04S3/00 (2006.01) i, H04S5/02 (2006.01) i			
According to International Patent Classification (IPC) or to both national classification and IPC			
B. FIELDS SEARCHED			
Minimum documentation searched (classification system followed by classification symbols)			
Int.Cl. G10L19/008, G10L19/00, H04S3/00, H04S5/02			
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched			
Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2014 Registered utility model specifications of Japan 1996-2014 Published registered utility model applications of Japan 1994-2014			
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)			
C. DOCUMENTS CONSIDERED TO BE RELEVANT			
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	
X	JP 7-264144 A (Toshiba Corporation) 1995.10.13, [0010] & US 5619197 A & EP 673167 A2 & KR 10-0192700 B	4-6	
X	JP 2009-278381 A (Japan Broadcasting Corporation) 2009.11.26, [0021]-[0023] (No family)	7-9	
A	JP 2010-521013 A (LG Electronics Incorporated) 2010.06.17, [0086]-[0088] & US 2010/0191354 A1 & EP 2137726 A & WO 2008/111773 A1 & KR 10-2008-0082924 A & CN 101675472 A & AT 526663 T & RU 2419168 C	7-9	
☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.			
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family			
Date of the actual completion of the international search		Date of mailing of the international search report	
12.02.2014		25.02.2014	
Name and mailing address of the ISA/JP		Authorized officer	
Japan Patent Office		Yoshitaka MAMIYA	
3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan		52 3454	
		Telephone No. +81-3-3581-1101 Ext. 3591	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/007390

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 9-50673 A (Sony Corporation) 1997.02.18, whole texts and figures & US 6112012 A & EP 757491 A2 & DE 69616450 D & CA 2182151 A & AT 208115 T & CN 1148237 A & MY 116545 A & CA 2182151 A1	4-6
A	JP 2005-49436 A (Toshiba Corporation) 2005.02.24, whole texts and figures & US 2005/0027522 A1 & US 2008/0091422 A1	4-6
A	JP 2011-35784 A (Sharp Corporation) 2011.02.17, whole texts and figures (No family)	7-9
A	JP 2008-154065 A (Roland Corporation) 2008.07.03, whole texts and figures (No family)	7-9
A	JP 2011-234177 A (Panasonic Corporation) 2011.11.17, whole texts and figures (No family)	7-9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/007390

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 1-3
because they relate to subject matter not required to be searched by this Authority, namely:
Claims 1-3 only comprise fields and information, although claims 1-3 denote that they are methods. This fact means that claims 1-3 are substantially [mere presentations of information] and are the subject matter which the International Searching Authority is not required to search under Article 17(2)(a)(i) and [Rule 39.1(v)].
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.