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

(11) **EP 4 178 228 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
10.05.2023 Bulletin 2023/19

(21) Application number: **21206983.5**

(22) Date of filing: **08.11.2021**

(51) International Patent Classification (IPC):
H04R 25/00 *(2006.01)*

(52) Cooperative Patent Classification (CPC):
H04R 25/558; H04R 2225/41

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(71) Applicant: **Sonova AG**
8712 Stäfa (CH)

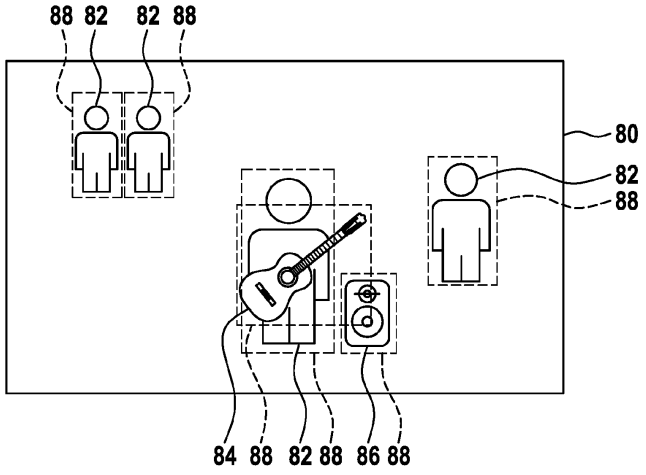
(72) Inventors:
- **HILDEBRAND, Nicola**
  8610 Uster (CH)
- **GRIEPENTROG, Sebastian**
  8712 Staefa (CH)
- **VON HOLTEN, Daniel**
  8713 Uerikon (CH)

(74) Representative: **Qip Patentanwälte**
Dr. Kuehn & Partner mbB
Goethestraße 8
80336 München (DE)

(54) **METHOD AND COMPUTER PROGRAM FOR OPERATING A HEARING SYSTEM, HEARING SYSTEM, AND COMPUTER-READABLE MEDIUM**

(57) A method for operating a hearing system (10) is provided. The hearing system (10) comprises a hearing device (12) configured to be worn at an ear of a user, a user device (14) communicatively coupled to the hearing device (12) and comprising a camera (36) and a display (30). The hearing device (12) comprises at least one sound input module (20) for generating an audio signal indicative of a sound detected in an environment of the hearing device (12), a first processing unit (40) for modifying the audio signal, and at least one sound output module (22) for outputting the modified audio signal. The method comprises: receiving image data from the camera (36), the image data being representative for a scene (80) in front of the camera (36); receiving an audio signal from the at least one sound input module (20), the audio signal being representative for the acoustic environment of the hearing device (12) substantially at a time the image data have been captured, wherein the acoustic environment comprises at least one audio source and wherein the audio signal is at least in part representative for a sound from the audio source; determining at least one visual object (88) as the audio source, within the scene (80) from the image data and the audio signal.

Fig. 6



EP 4 178 228 A1

Description

FIELD OF THE INVENTION

[0001] The invention relates to a method and a computer program for operating a hearing system, to the hearing system, and to a computer-readable medium in which the computer program is stored.

BACKGROUND OF THE INVENTION

[0002] Hearing devices are generally small and complex devices. A typical hearing device comprises a processing unit, e.g. including one or more processors, a sound input module, e.g. a microphone, a sound output module, e.g. an loudspeaker, a memory communicatively coupled to the processing unit, a housing, and other electronical and mechanical components. Some example hearing devices are Behind-The-Ear (BTE), Receiver-In-Canal (RIC), In-The-Ear (ITE), Completely-In-Canal (CIC), and Invisible-In-The-Canal (IIC) devices. A user can prefer one of these hearing devices compared to another device based on hearing loss, aesthetic preferences, lifestyle needs, and budget.

[0003] Sometimes when hearing device users use their hearing device in everyday life, they may notice that the hearing device does not always support them well enough. Be it that the hearing device setting is not correctly set for the current acoustic environment, e.g. with respect to audio sources in the environment and/or a listening activity of the user, or an automatic classifier does not classify the situation correctly based on an acoustic input only.

[0004] The options, that a user conventionally has, are modification possibilities that are named for the acoustic intention, e.g. reduce background noise, low, high etc., or describe technical possibilities. Another option is to switch to an appropriate manual program if the automatic mode does not classify the acoustic environment correctly.

[0005] The translation of hearing intentions or hearing problems with audio sources in complex listening situations into adjustment actions of the modifier is challenging and sometimes impossible for the hearing device user. The range of modifiers that can be used to adjust the hearing device according to the listening intention might be overwhelming. The uncertainty, whether the right modifier is used to improve the performance is high and the risk is high that users adjust not the correct hearing device behavioural part.

[0006] Fast and efficient adjustments done by the user may be crucial and a match decision for a future engagement of the user to use the modification/optimization functionality offered may afford the user to react quickly to changes in the acoustic environment.

DESCRIPTION OF THE INVENTION

[0007] It is an objective of the present invention to provide a method and a computer program for operating a hearing system, which enables a quick and/or easy in-the-field-setting of the hearing system for a user of the hearing system, in particular in complex acoustic environments. It is another objective of the present invention to provide the hearing system and a computer-readable medium in which the computer program is stored.

[0008] These objectives are achieved by the subject-matter of the independent claims. Further exemplary embodiments are evident from the dependent claims and the following description.

[0009] A first aspect relates to a method for operating a hearing system. The hearing system comprises a hearing device configured to be worn at an ear of a user, a user device communicatively coupled to the hearing device and comprising a camera and a display, with the hearing device comprising at least one sound input module for generating an audio signal indicative of a sound detected in an environment of the hearing device, a first processing unit for modifying the audio signal, and at least one sound output module for outputting the modified audio signal. The method comprises: receiving image data from the camera, the image data being representative of a scene in front of the camera; receiving an audio signal from the at least one sound input module, the audio signal being representative of the acoustic environment of the hearing device substantially at a time the image data have been captured, wherein the acoustic environment comprises at least one audio source and wherein the audio signal is at least in part representative for a sound from the audio source; and determining at least one visual object as the audio source, within the scene from the image data and the audio signal.

[0010] The method may be a computer-implemented method, which may be performed automatically by the hearing system. The step of determining the at least one visual object as the audio source within the scene from the image data and the audio signal may be carried out by an artificial intelligence and/or a neural network. The artificial intelligence or, respectively the neural network, may be trained with the data set comprising a huge amount of image data representing different scenes with visual objects, wherein at least some of the visual objects are the audio sources, and a corresponding amount of audio signals associated and/or synchronized with the image data.

[0011] The hearing system may, for instance, comprise one or two hearing devices used by the same user. One or both of the hearing devices may be worn on or in an ear of the user. A hearing device may be a hearing aid, which may be adapted for compensating a hearing loss of the user. Also, a cochlear implant may be a hearing device or at least a part of it. The hearing system may optionally further comprise at least one connected user device, such as a smartphone, smartwatch, smart glass-

es, or another device carried by the user or a personal computer of the user etc. The visual objects may be determined from the image data by analysing the image data with the help of a data base comprising several different objects and/or object classes. The time the image data have been captured may be encoded in meta data accompanying the image data. Alternatively or additionally, the image data may be captured and received in real time such that the time of receiving the image data automatically corresponds to the time the audio signal is captured and received. The visual object may be selected by a touch on the display, if the display is a touch screen.

[0012] As explained above, the audio signal is representative of the acoustic environment of the hearing device substantially at a time the image data have been captured, wherein "substantially" may mean in this context, that the audio signal is representative of the acoustic environment of the hearing device at the time the image data have been captured, that the audio signal is representative of the acoustic environment of the hearing device during a time interval in which the image data have been captured, or that the audio signal is representative of the acoustic environment of the hearing device during a time interval which overlaps a time interval during which the image data have been captured. The audio signal may comprise meta data representing the time or time interval during which the corresponding sound of the acoustic environment has been captured. Alternatively or additionally, the image may comprise meta data representing the time or time interval during which the image data have been captured. The audio signal and the image data may be synchronized by these meta data.

[0013] A second aspect relates to the hearing system. The hearing system comprises: the hearing device configured to be worn at an ear of a user and comprising the at least one sound input module for generating the audio signal, the first processing unit for modifying the audio signal, and the at least one sound output module for outputting the modified audio signal; and the user device communicatively coupled to the hearing device and comprising the display, the camera, and a second processing unit; wherein at least one control unit is coupled to the hearing device and the user device and is configured to carry out the above method. The hearing system may further include, by way of example, a second hearing device worn by the same user. If the user device is a smartphone, the camera may be the camera implemented within the smartphone. Alternatively, the camera may be implemented in smart glasses, if the user device is the smart glasses.

[0014] A third aspect relates to a computer program for operating the hearing system, which program, when being executed by a processing unit, e.g. the first and/or second processing unit, is adapted to carry out the steps of the above method.

[0015] A fourth aspect relates to a computer-readable medium, in which the above computer program is stored. In general, a computer-readable medium may be a floppy disk, a hard disk, an USB (Universal Serial Bus) storage device, a RAM (Random Access Memory), a ROM (Read Only Memory), an EPROM (Erasable Programmable Read Only Memory) or a FLASH memory. A computer-readable medium may also be a data communication network, e.g. the Internet, which allows downloading a program code. The computer-readable medium may be a non-transitory or transitory medium.

[0016] For example, the computer program may be executed in the first processing unit of the hearing device, which hearing device, for example, may be carried by the person behind the ear. The computer-readable medium may be a memory of this hearing device. The computer program also may be executed by the second processing unit of the connected user device, such as a smartphone or any other type of mobile device, which may be a part of the hearing system, and the computer-readable medium may be a memory of the connected user device. It also may be that some steps of the method are performed by the hearing device and other steps of the method are performed by the connected user device.

[0017] It has to be understood that features of the method as described above and in the following may be features of the computer program, the computer-readable medium and/or the hearing system as described above and in the following, and vice versa.

[0018] Determining the at least one visual object as the audio source within the scene from the image data and the audio signal enable a quick and/or easy in-the-field-setting of the hearing system for the user of the hearing system, in particular in complex acoustic environments. In general, in-the-field-setting is getting more and more important as it has many advantages compared with the conventional hearing device setting in a sound booth. The above method contributes to increasing the trust of a customer as it ensures a quick and easy solution for difficult listening situations, i.e. complex acoustic environments with e.g. several audio and/or noise sources.

[0019] According to an embodiment of the invention, the method further comprises: displaying the scene on the display and marking the determined visual object within the scene; receiving an input of the user, the input being representative for the user selecting the marked visual object and for the user wishing to selectively modify the sound from the audio source associated with the selected visual object; selectively modifying the audio signal from the audio source associated with the selected visual object; and outputting the modified audio signal to the user. This gives the user the possibility to visually selecting the corresponding sound sources. With this in-the-field-setting approach of the above method, more individual user data may be obtained. Also, users without a good technical understanding can use that approach very well. Especially data regarding the users' individual visual selection of the visual objects and the associated audio source of interest, what corresponds to a listening intention of the user, may be valuable for an AI-based setting.

[0020] According to an embodiment, if the step of determining at least one visual object as the audio source, within the scene from the image data and the audio signal, is not carried out by an artificial intelligence and/or a neural network, the step of determining at least one visual object as the audio source, within the scene from the image data and the audio signal may comprise: determining the at least one visual object, which is a potential audio source, within the scene from the image data; determining the at least one audio source within the acoustic environment from the audio signal; and associating at least one determined visual object with at least one determined audio source, wherein the determined visual object may be associated with the determined audio source. For example, several visual objects are determined and several audio sources are determined. Then, some of the several determined visual objects may be associated with one of the several determined audio sources each.

[0021] According to an embodiment, the step of determining the at least one visual object, which is the potential audio source, comprises determining a first spatial relationship between the camera and the visual object, e.g. a direction and/or a distance from the camera to the visual object; the step of determining the at least one audio source within the acoustic environment comprises determining a second spatial relationship between the hearing device and the audio source, e.g. a direction and/or a distance from the hearing device to the audio source, wherein the audio signal may be a stereo-signal; and the step of associating the at least one determined visual object with the at least one determined audio source comprises comparing the first spatial relationships of all determined visual objects with the second spatial relationships of all determined sound sources, and associating that visual object with that audio source such that the corresponding first and second spatial relationship fulfil a predetermined requirement. The predetermined requirement may be the "best fit" of the spatial relationships between the camera and the visual object and the hearing device and the audio source.

[0022] According to an embodiment, the method further comprises classifying the acoustic environment from the received audio signal; modifying the audio signal in accordance with the classification; and determining the at least one audio source within the acoustic environment from the modified audio signal. Depending on the classified acoustic environment, a set of feature parameters is selected as a determined sound program. With such an acoustic environment classification, an acoustic situation the wearer is in is classified and consequently categorized in order to automatically adjust the features and/or values of parameters of these features in accordance with the current acoustic situation. Optionally a feature activity may be logged such that it is logged which feature is active at which time. Classifying the acoustic environment and modifying the audio signal in accordance with the classification by the corresponding sound program including a set of features contributes to identify a signal to noise ratio, an object loudness, a type of noise or room acoustic, a pitch of the object or other information which may be required to choose the most effective sound cleaner or frequency dependent gain modification.

[0023] According to an embodiment, the method further comprises determining at least one object class of the determined visual object and labelling the marked visual object in the scene in accordance with the determined object class. The labels assist the user in identifying the marked visual object. The object class may be at least one of the group of human, animal, instrument, speaker, dishes, newspaper, car, and water. For each object class, auto-adjustments, macro modifications and/or a list of modifiers may be defined, which may have an impact on the sound from the corresponding sound source.

[0024] According to an embodiment, the method further comprises providing at least one input field on the display for the input of the user, with the input field being representative for at least one modification of the audio signal with respect to the sound from the audio source assigned to the selected visual object, wherein the audio signal is selectively modified with respect to the sound from the audio source assigned to the selected visual object, if the user activates the input field. The input field may be activated by a direct pressure on the input field or by a gesture above, on or next to the input field. The input field on the display represents an intuitive possibility for the user to quickly and easily set the preferred modification.

[0025] According to an embodiment, the input field is provided depending on the classification of the acoustic environment, with the input field being representative for at least one modification in accordance with the classification of the acoustic environment. Alternatively or additionally, the input field is provided depending on the object class of the determined visual object, with the input field being representative for at least one modification in accordance with the object class of the determined visual object. Providing the input field depending on the classification enables to provide the user with the optimal and/or preferred option for modifying the audio signal in the current acoustic situation.

[0026] According to an embodiment, the input is representative for the user wishing to increase the volume of the sound from the selected visual object and/or to decrease the volume or effectuate a dampening of the sound from all other determined visual objects, or to decrease the volume or effectuate a dampening of the sound from the selected visual object and/or to increase the volume of the sound from all other determined visual objects, and the audio signal is selectively modified such that the volume of the audio source associated with the selected visual object is increased or, respectively decreased, or that the volume of the audio sources of all other determined visual objects is decreased or, respectively, increased.

[0027] According to an embodiment, the method further comprises monitoring the visual object by the camera and stopping to selectively modify the audio signal with respect to the sound from the audio source assigned to the monitored visual object if the visual object disappears from a field of view of the camera. This enables to save processing resources for selectively modifying the audio signal with respect to the sound from the audio source assigned to the monitored visual object, if the visual object and as such the audio source disappears. The audio signal from the visual object may be modified again as soon as the visual object is recognized within the scene again.

[0028] According to an embodiment, the method further comprises detecting at least one gesture of the user on or above the display; and selecting the marked visual object in accordance with the gesture; and/or selectively modifying the audio signal with respect to the sound from the audio source associated with the selected visual object in accordance with the gesture. Detecting the gesture provides a very intuitive input possibility for the user.

[0029] According to an embodiment, the hearing system further comprises a remote server communicatively coupled to the hearing device and/or the user device and being configured to carry out at least a part of the above method. The provision of the server enables to outsource processing tasks from the hearing device and/or the user device to the server. This is especially advantageous, if the corresponding processing tasks need huge processing resources and/or if the hearing device and, respectively, the user device have to be relieved.

[0030] According to an embodiment, the control unit is implemented in the first processing unit, the second processing unit or the remote server. In other words, the processing of the above method may be controlled by the hearing device, the user device, or, respectively, the remote server. The user device may be connected to a cloud and/or the internet. Some of the steps of the method described here and further below may be executed on the hearing device, the user device or in the cloud, or any combination thereof.

[0031] These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments described hereinafter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0032] Below, embodiments of the present invention are described in more detail with reference to the attached drawings.

Fig. 1 shows a hearing system according to an embodiment of the invention.

Fig. 2 shows a block diagram of components of the hearing system according to figure 1.

Fig. 3 shows a flow diagram of a method for operating a hearing system according to an embodiment of the invention.

Fig. 4 shows a flow diagram of a sub-method of the method of figure 3.

Fig. 5 shows an example of a scene including several visual objects.

Fig. 6 shows the scene of figure 5 with several marked visual objects.

Fig. 7 shows the scene of figure 6 with two selected visual objects.

Fig. 8 shows another example of a scene including several marked and labelled visual objects.

Fig. 9 shows the scene of figure 5 and two examples of input fields.

Fig. 10 shows another example of a scene including a marked and labelled visual object and several examples for selecting a proper modification of the corresponding audio source.

Fig. 11 shows an example of a visual object-based audio signal modifier implementation.

[0033] The reference symbols used in the drawings, and their meanings, are listed in summary form in the list of reference symbols. In principle, identical parts are provided with the same reference symbols in the figures.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0034] Fig. 1 schematically shows a hearing system 10 according to an embodiment of the invention. The hearing system 10 includes a hearing device 12 and a user device 14 connected to the hearing device 12. As an example, the hearing device 12 is formed as a behind-the-ear device carried by a user (not shown) of the hearing device 12. It has to be noted that the hearing device 12 is a specific embodiment and that the method described herein also may be performed with other types of hearing devices, such as e.g. an in-the-ear device or one or two of the hearing devices 12 mentioned above. The user device 14 may be a smartphone, a tablet computer, and/or smart glasses.

[0035] The hearing device 12 comprises a part 15 behind the ear and a part 16 to be put in the ear channel of the user. The part 15 and the part 16 are connected by a tube 18. The part 15 comprises at least one sound input module 20, e.g. a microphone or a microphone array, a sound output module 22, such as a loudspeaker, and an input mean 24, e.g. a knob, a button, or a touch-sensitive sensor, e.g. capacitive sensor. The sound input module

20 can detect a sound in the environment of the user and generate an audio signal indicative of the detected sound. The sound output module 22 can output sound based on the audio signal modified by the hearing device 12, wherein the sound from the sound output module 22 is guided through the tube 18 to the part 16. The input mean 24 enables an input of the user into the hearing device 12, e.g. in order to power the hearing device 12 on or off, and/or for choosing a sound program or any other modification of the audio signal.

[0036] The user device 14 comprises a display 30, e.g. a touch-sensitive display, providing a graphical user interface 32 including control element 32, e.g. a slider, which may be controlled via a touch on the display 30, and a camera 36. The control element 32 may be referred to as input means of the user device 14. The camera 36 may be a photo camera and/or video camera. If the user device 14 is the smart glasses, the use device 14 may comprise a knob or button instead of the display 30 and/or the graphical user interface 32.

[0037] Fig. 2 shows a block diagram of components of the hearing system 10 according to figure 1.

[0038] The hearing device 12 comprises a first processing unit 40. The first processing unit 40 is configured to receive the audio signal generated by the sound input module 20. The hearing device 12 may include a sound processing module 42. For instance, the sound processing module 42 may be implemented as a computer program executed by the first processing unit 40. The sound processing module 42 may be configured to modify, in particular increase or decrease a volume of and/or delay, the audio signal generated by the sound input module 20, e.g. some frequencies or frequency ranges of the audio signal depending on parameter values of parameters, which influence the amplification, the damping and/or, respectively, the delay, e.g. in correspondence with a current sound program. The parameter may be one or more of the group of frequency dependent gain, time constant for attack and release times of compressive gain, time constant for noise canceller, time constant for dereverberation algorithms, reverberation compensation, frequency dependent reverberation compensation, mixing ratio of channels, gain compression, gain shape/amplification scheme. A set of one or more of these parameters and parameter values may correspond to a predetermined sound program, wherein different sound programs are characterized by correspondingly different parameters and parameter values. The sound program may comprise a list of sound processing features. The sound processing features may for example be a noise cancelling algorithm or a beamformer, which strengths can be increased to increase speech intelligibility but with the cost of more and stronger processing artifacts. The sound output module 22 generates sound from the modified audio signal and the sound is guided through the tube 18 and the in-the-ear part 16 into the ear channel of the user.

[0039] The hearing device 12 may include a control module 44. For instance, the control module 44 may be implemented as a computer program executed by the first processing unit 40. The control module 44 may be configured for adjusting the parameters of the sound processing module 42, e.g. such that an output volume of the sound signal is adjusted based on an input volume. For example, the user may select a modifier (such as bass, treble, noise suppression, dynamic volume, etc.) and levels and/or values of the modifiers with the input mean 24. From this modifier, an adjustment command may be created and processed as described above and below. In particular, processing parameters may be determined based on the adjustment command and based on this, for example, the frequency dependent gain and the dynamic volume of the sound processing module 42 may be changed.

[0040] All these functions may be implemented as different sound programs stored in a first memory 50 of the hearing device 12, which sound programs may be executed by the sound processing module 42. The first memory 50 may be implemented by any suitable type of storage medium, in particular a non-transitory computer-readable medium, and can be configured to maintain, e.g. store, data controlled by the first processing unit 40, in particular data generated, accessed, modified and/or otherwise used by the first processing unit 40. The first memory 50 may also be configured to store instructions for operating the hearing device 12 and/or the user device 14 that can be executed by the first processing unit 40, in particular an algorithm and/or a software that can be accessed and executed by the first processing unit 40.

[0041] A sound source detector 46 may be implemented in a computer program executed by the first processing unit 40. The sound source detector 46 is configured to determine at least the one sound source from the audio signal. In particular, the sound source detector 46 may be configured to determine a spatial relationship between the hearing device 12 and the corresponding sound source. The spatial relationship may be given by a direction and/or a distance from the hearing device 12 to the corresponding audio source, wherein the audio signal may be a stereo-signal and the direction and/or distance may be determined by different arrival times of the sound waves from one audio source at two different sound input modules 20 of the hearing device 12 and/or a second hearing device 12 worn by the same user.

[0042] A first classifier 48 may be implemented in a computer program executed by the first processing unit 40. The first classifier 48 can be configured to evaluate the audio signal generated by the sound input module 20. The first classifier 48 may be configured to classify the audio signal generated by the sound input module 20 by assigning the audio signal to a class from a plurality of predetermined classes. The first classifier 48 may be configured to determine a characteristic of the audio signal generated by the sound input module 20, wherein the audio signal is assigned to the class depending on the determined characteristic. For instance, the first classifier

48 may be configured to identify one or more predetermined classification values based on the audio signal from the sound input module 20. The classification may be based on a statistical evaluation of the audio signal and/or a machine learning (ML) algorithm that has been trained to classify the ambient sound, e.g. by a training set comprising a huge amount of audio signals and associated classes of the corresponding acoustic environment. So, the ML-algorithm may be trained with several audio signals of acoustic environments, wherein the corresponding classification is known.

[0043] The first classifier 48 may be configured to identify at least one signal feature in the audio signal generated by the sound input module 20, wherein the characteristic determined from the audio signal corresponds to a presence and/or absence of the signal feature. Exemplary characteristics include, but are not limited to, a mean-squared signal power, a standard deviation of a signal envelope, a mel-frequency cepstrum (MFC), a mel-frequency cepstrum coefficient (MFCC), a delta mel-frequency cepstrum coefficient (delta MFCC), a spectral centroid such as a power spectrum centroid, a standard deviation of the centroid, a spectral entropy such as a power spectrum entropy, a zero crossing rate (ZCR), a standard deviation of the ZCR, a broadband envelope correlation lag and/or peak, and a four-band envelope correlation lag and/or peak. For example, the first classifier 48 may determine the characteristic from the audio signal using one or more algorithms that identify and/or use zero crossing rates, amplitude histograms, auto correlation functions, spectral analysis, amplitude modulation spectrums, spectral centroids, slopes, roll-offs, auto correlation functions, and/or the like. In some instances, the characteristic determined from the audio signal is characteristic of an ambient noise in an environment of the user, e.g. a noise level, and/or a speech, e.g. a speech level. The first classifier 48 may be configured to divide the audio signal into a number of segments and to determine the characteristic from a particular segment, e.g. by extracting at least one signal feature from the segment. The extracted feature may be processed to assign the audio signal to the corresponding class.

[0044] The first classifier 48 may be further configured to assign, depending on the determined characteristic, the audio signal generated by the sound input module 20 to a class of at least two predetermined classes. The classes may represent a specific content in the audio signal. For instance, the classes may relate to a speaking activity of the user and/or another person and/or an acoustic environment of the user. Exemplary classes include, but are not limited to, low ambient noise, high ambient noise, traffic noise, music, machine noise, babble noise, public area noise, background noise, speech, non-speech, speech in quiet, speech in babble, speech in noise, speech in loud noise, speech from the user, speech from a significant other, background speech, speech from multiple sources, calm situation and/or the like. The first classifier 48 may be configured to evaluate the characteristic relative to a threshold. The classes may comprise a first class assigned to the audio signal when the characteristic is determined to be above the threshold, and a second class assigned to the audio signal when the characteristic is determined to be below the threshold. For example, when the characteristic determined from the audio signal corresponds to ambient noise, a first class representative of a high ambient noise may be assigned to the audio signal when the characteristic is above the threshold, and a second class representative of a low ambient noise may be assigned to the audio signal when the characteristic is below the threshold. As another example, when the characteristic determined from the audio signal is characteristic of a speech, a first class representative of a larger speech content may be assigned to the audio signal when the characteristic is above the threshold, and a second class representative of a smaller speech content may be assigned to the audio signal when the characteristic is below the threshold.

[0045] At least two of the classes can be associated with different sound programs, in particular with different sound processing parameters, which may be applied by the sound processing module 42 for modifying the audio signal. To this end, the class assigned to the audio signal, which may correspond to a classification value, may be provided to the control module 44 in order to select the associated audio processing parameters, in particular the associated sound program, which may be stored in the first memory 50. The class assigned to the audio signal may thus be used to determine the sound program, which may be automatically used by the hearing device 12, in particular depending on the audio signal received from the sound input module 20.

[0046] The hearing device 12 may further comprise a first transceiver 52. The first transceiver 52 may be configured for a wireless data communication with a remote server 72. Additionally or alternatively, the first transceiver 52 may be adapted for a wireless data communication with a second transceiver 64 of the user device 14. The first and/or the second transceiver 52, 64 each may be e.g. a Bluetooth or RFID radio chip.

[0047] Each of the sound processing module 42, the control module 44, the sound source detector 46, and the first classifier 48 may be embodied in hardware or software, or in a combination of hardware and software. Further, at least two of the modules 42, 44, 46, 48 may be consolidated in one single module or may be provided as separate modules. The first processing unit 40 may be implemented as a single processor or as a plurality of processors. For instance, the first processing unit 40 may comprise a first processor in which the sound processing module 42 is implemented, and a second processor in which the control module 44 and/or the sound source detector 46 and/or the first classifier 48 are implemented.

[0048] The user device 14, which may be connected to the hearing device 12, may comprise a second processing unit 60, a second memory 62, a second trans-

ceiver 64, a second classifier 66 and/or a visual object detector 68.

[0049] The second processing unit 60 may comprise one or more processors, e.g. for running the second classifier 66 and/or the visual object detector 68. If the hearing device 12 is controlled via the user device 14, the second processing unit 60 of the user device 14 may be seen at least in part as a controller of the hearing device 12. In other words, according to some embodiments, the first processing unit 40 of the hearing device 12 and the second processing unit 60 of the user device 14 may form the controller of the hearing device 12. A processing unit of the hearing system 10 may comprise the first processing unit 40 and the second processing unit 60. The processing units 40, 60 may communicate data via the first and second transceivers 52, 64.

[0050] The second classifier 66 may have the same functionality as the first classifier 48 explained above. The second classifier 66 may be arranged alternatively or additionally to the first classifier 48 of the hearing device 12. The second classifier 66 may be configured to classify the acoustic environment of the user and the user device 14 depending on the received audio signal, as explained above with respect to the first classifier 48, wherein the acoustic environment of the user and the user device 14 corresponds to the acoustic environment of the hearing device 12 and wherein the audio signal may be forwarded from the hearing device 12 to the user device 14.

[0051] The visual object detector 68 is configured to identify one or more visual objects 88 (see figure 4) within the scene 80 taken by the camera 36. Further, the visual object detector 68 may be configured to communicate with a display control (not shown) of the user device of 14, such that the display control controls the display 30 in order to visually mark the determined visual objects 88. The visual object detector 68 may be implemented as an algorithm and/or an artificial intelligence. In case of an artificial intelligence, the algorithm has to be trained, e.g. with the help of the huge set of image data representing pictures each including one or more visual objects. A list of potential visual objects may be stored in a database, which may be stored in the first and/or second memory 50, 62. Further, a set of adjustable audio sources and corresponding object classes may be predefined, like e.g. human, instrument, speaker, dishes, newspaper, car, water, noise, etc. For each of these audio sources, auto-adjustments, macro modifications and/or a list of modifiers may be predefined, which may have an impact on the corresponding sound object. An exemplary auto-adjustment may be decreasing an overall gain, e.g. of 1.5dB, increasing a noise reduction strength, e.g. to 0.8, and increasing a beamformer strength, e.g. to 0.85.

[0052] There are several conventional high-performing algorithms available, which are able to detect visual objects within an image. One of the best performing algorithms is the YOLO (You Only Look Once) method. This algorithm is able to run on a smartphone in real time as it processes only one image at the same time. Hence, it is proposed to use the image data of the camera 36 or alternatively from a smart glasses camera, a smartwatch camera, a TV camera, or a recorded video, to run the visual object detection algorithm on the user device 14.

[0053] With the hearing system 10 it is possible that the above-mentioned modifiers and their levels and/or values are adjusted with the user device 14 and/or that the adjustment command is generated with the user device 14. This may be performed with a computer program run in the second processing unit 60 and stored in the second memory 62 of the user device 14. This computer program may also provide the graphical user interface 32 on the display 30 of the connected user device 14. For example, for adjusting the modifier, such as volume, the graphical user interface 32 may comprise the control element 34, such as a slider. When the user adjusts the slider, an adjustment command may be generated, which will change the sound processing of the hearing device 12 as described above and below. Alternatively or additionally, the user may adjust the modifier with the hearing device 12 itself, for example via the input mean 24.

[0054] The hearing device 12 and/or the user device 14 may communicate with each other and/or with the remote server 72 via the Internet 70. The method explained below with respect to figures 3 and/or 4 may be carried out at least in part by the remote server 72. For example, processing tasks, which require a huge amount of processing resources, may be outsourced from the hearing device 12 and/or the user device of 14 to the remote server 72. For example, the determination of the visual objects from the image data and/or of the audio sources from the audio signal, may be outsourced to the remote server 72. Further, the processing units (not shown) of the remote server 72 may be used at least in part as the controller for controlling the hearing device 12 and/or the use device 14

[0055] Fig. 3 shows a flow diagram of a method for operating the hearing system 10. The method may be carried out by the first and/or the second processing unit 40, 60 and/or by the remote server 72, wherein some of the steps of the method may be carried out by the first and/or the second processing unit 40, 60 and/or some other steps of the method may be carried out by the remote server 72.

[0056] In a step S2, image data from the camera 36 are received, e.g. by the first or second processing unit 40. The image data are representative for a scene 80 in front of the user. In particular, a picture or video of the front of the user may be taken by the camera 36 and the camera 36 generates the image data representing the scene 80 in front of the user.

[0057] In a step S4, an audio signal from the at least one sound input module 20 may be received, e.g. by the first or second processing unit 40. The audio signal is representative for the acoustic environment of the user at a time the image data have been captured. The acoustic environment comprises at least one audio source and

the audio signal is at least in part representative for a sound from the audio source.

[0058] In a step S6, at least one visual object 88 is determined as the audio source, within the scene 80 from the image data and the audio signal. Step S6 may be carried out by an artificial intelligence and/or by a "traditional" algorithm.

[0059] In a step S8, the scene 80 is displayed on the display 30.

[0060] Fig. 5 shows an example of the scene 80 including several visual objects 88. In particular, the scene 80 comprises four persons 82, wherein two of the persons 82 are seen in the upper left in the background of the scene 80, one person 82 is shown at the right in the background of the scene 80 and one person 82 is shown in the middle in the front of the scene 80. Further, the scene 80 comprises an instrument 84 and a speaker 86 both representing visual objects 88 within the scene 80 and potential audio sources.

[0061] In a step S10, at least one determined visual object 88 is marked within the scene 80. Further, the user may be prompted to select at least one marked visual object 88 within the scene 80. If one of the visual objects 88 is detected as an adjustable audio source, the user has to become aware of it. One way to indicate a detected audio source is to highlight the corresponding visual object 88 within the scene 80. This may be done with augmenting a boundary around the detected visual object 88 within the scene 80. Another way to indicate the detected visual object 88 acting as an audio source is to show a corresponding object related modifier directly within the scene 80 (see figures 9 and/or 10).

[0062] Fig. 6 shows the scene 80 of figure 5 with several marked visual objects 88. In particular, the persons 82, the instrument 84 and the speaker 86 are identified and marked as visual objects 88 potentially acting as audio sources. The visual objects 88 are marked by the fine dashed rectangles, as an example.

[0063] In a step S12, an input of the user is received. The input is representative for that the user selected at least one of the marked visual objects 88. Each marked visual object 88, which has been selected by the user, is referred to as selected visual object 90 in the following. The input is further representative for that the user wishes to selectively modify the sound from the audio source associated with the selected visual object 90. Optionally, the input may be representative for the user wishing to increase the volume of the sound from the selected visual object 90 and/or to decrease the volume or effectuate a dampening of the sound from all other determined visual objects 88. Alternatively, the input may be representative for the user wishing to decrease the volume or effectuate a dampening of the sound from the selected visual object 90 and/or to increase the volume of the sound from all other determined visual objects 88. If the visual objects 88 potentially acting as an audio source are highlighted within the scene 80, e.g. with an augmented boundary around the corresponding visual object 80, the augmented boundary itself may be a hitbox to select the corresponding visual object 80. If the object-based modifier is shown directly when a visual object 80 acting as an audio source is detected within the scene 80, this step becomes obsolete, because the user automatically selects one of the visual objects 80, if he/she activates the corresponding object-based modifier.

[0064] Fig. 7 shows the scene of figure 6 with two selected visual objects 90. In particular, in figure 7, the instrument 84 and the speaker 86 are marked as selected visual objects 90, e.g. by course dashed rectangles.

[0065] In a step S14, the audio signal from the audio source associated with the selected visual object 90 is selectively modified. Optionally, the audio signal may be selectively modified such that the audio source associated with the selected visual object 90 is amplified or that the volume of the audio sources of all other determined visual objects 88 is decreased or a dampening thereof is effectuated. Alternatively the audio signal may be selectively modified such that the volume of the audio source associated with the selected visual object 90 is decreased or a dampening thereof is effectuated, and/or that the volume of the audio sources of all other determined visual objects 88 is increased.

[0066] In a step S16, the modified audio signal is outputted to the user.

[0067] Optionally, the method further comprises the step of monitoring the visual object(s) 88 by the camera 36 and stopping to selectively modify the audio signal with respect to the sound from the audio source assigned to the monitored visual object 88, if the visual object 88 disappears from a field of view of the camera 36.

[0068] A pure object detection might not be sufficient for some applications as it is not able to detect the signal to noise ratio, object loudness, type of noise or room acoustic, pitch of the object or other information which may be required to choose the most effective sound cleaner or frequency dependent gain modification. Therefore, the above method may further comprise the steps of classifying the acoustic environment from the received audio signal, modifying the audio signal in accordance with the corresponding classification, and step S6, e.g. determining the at least one audio source within the acoustic environment, may be carried out using the modified audio signal.

[0069] A combination of a visual and an acoustic classification may be even better able to provide the optimal adjustment parameter. Therefore, the above method may further comprise the steps of determining at least one object class of the determined visual object 88. A list of adjustable visual objects acting as audio sources or corresponding sound object classes may be available in a corresponding database stored within the first and/or second memory 50, 62 and/or within the remote server 72. The corresponding visual and acoustic object detection algorithm may be trained to estimate the most possible and effective sound object class and its adjustment suggestion. This adjustment suggestion may be applied

automatically or for manual adjustments offered to the user.

[0070] Fig. 4 shows a flow diagram of a sub-method of the method of figure 3. In particular, figure 4 shows a sub-routine of the above method in case the method is implemented by a "traditional" algorithm and not by an artificial intelligence and/or a neuronal network. In this case step S6, i.e. the step of determining at least one visual object 88 as the audio source within the scene 80 from the image data and the audio signal, may comprise the following steps S18, S20 and S22.

[0071] In step S18 at least one visual object 88, which is a potential audio source, is determined within the scene 80 from the image data. For example, step S18 comprises determining a first spatial relationship between the camera 36 and the visual object 88.

[0072] In step S20, at least one audio source is determined within the acoustic environment from the audio signal. For example, step S20 comprises determining a second spatial relationship between the hearing device 12 and the audio source.

[0073] In step S22, at least one determined visual object 88 is associated with at least one determined audio source. For example, the first spatial relationships of all determined visual objects 88 may be compared with the second spatial relationships of all determined sound sources. Then, that visual object 88 of all determined visual objects 88 may be associated with that audio source of all determined sound sources such that the corresponding first and second spatial relationship fulfil a predetermined requirement. The predetermined requirement may be the "best fit" of the spatial relationship between the camera and the visual object and the hearing device and the audio source.

[0074] Fig. 8 shows another example of a scene 80 including several marked and labelled visual objects 88 in accordance with the present invention, in particular in accordance with the above method. In particular, all marked visual objects 88 within the scene 80 are labelled in accordance with the determined object class. For example, the woman in the background in the upper left has the label 92 "Female", the man in the middle in the front has also the label "Male", the man in the background on the right has the label "Male", and the clapping hands of the man in the background on the right have the label "Clapping". The labelling may be carried out within steps S8 or S10 of the above method.

[0075] Further, one, two or more state indicators 94 may be shown within the scene 80. For example, the state indicators 94 may indicate the classification of the acoustic environment, if the acoustic environment is classified. In particular, the state indicators 94 may be representative for the noisy acoustic environment ("Noisy" in figure 8) and/or that the acoustic environment is within a room or house (house-symbol in figure 8).

[0076] Optionally, the above method further comprises the step of providing the at least one input field 96 on the display 30 for the input of the user such that the input field 96 is representative for at least one modification of the audio signal with respect to the sound from the audio source assigned to the selected visual object 90, wherein the audio signal is selectively modified with respect to the sound from the audio source assigned to the selected visual object 90, if the user activates the input field 96. For example, an input field 96 may be provided on the display 30 within the scene 80. The input field 96 may comprise a one, two or more, e.g. four, buttons 98. The user may use of the input field 96 and in particular the buttons 98 for the user input of step S12 of the above method. For example, some of the buttons 98 may be representative for that the proper visual object 88 is selected or not, e.g. such that the user can confirm or deny that the proper visual object 88 is selected. Alternatively or additionally, some of the buttons 98 may be representative for the predetermined modification of the audio signal with respect to the audio source of the selected visual object 90. For example, the user may indicate that he wishes to increase or decrease the volume of the sound from the selected visual object 90 by pressing the corresponding button 98.

[0077] Further, the input field 96 and in particular the buttons 98 may represent object-based modifiers. In general, an object-based modifier may be an input field 96 and/or a button 98, which appearance and/or function depends on the object class of the selected visual object 90. For example, in a first case some of the buttons 98 are representative for increasing or decreasing the volume of the sound of the selected visual object 90, if the selected visual object 90 is the man in the front of the scene 80, and in contrast the buttons 98 at the same position within the input field 96 may be representative for increasing or decreasing the beamformer strength, if the selected visual object 90 is the clapping hands.

[0078] Optionally, an overview of the detected visual objects 88 potentially acting as an audio source and/or the corresponding recommended sound adjustments, which may be performed in the background of a sound object-based modifier (see figure 11) may be shown to the user on the display 30, e.g. within the scene 80. In this way, the user can be instructed how to self-adjust the sound of the audio source corresponding to certain visual objects 88 properly.

[0079] Fig. 9 shows the scene of figure 5 and two examples of input fields 96, wherein the input fields 96 of figure 9 may represent object-based modifiers. For example, the input fields 96 each comprise a label 92 and two buttons 98 for increasing and, respectively decreasing, the value of one of the parameters for modifying the audio signal with respect to the sound from the audio source.

[0080] Optionally, one or more of the above input fields 96 are provided depending on the classification of the acoustic environment, with the corresponding input field 96 being representative for at least one modification in accordance with the classification of the acoustic environment. Alternatively or additionally, the corresponding

input field 96 may be provided depending on the object class of the determined visual object 88, with the input field 96 being representative for at least one modification in accordance with the object class of the determined visual object 88, wherein in this case the input field 96 may be also referred to as object-based modifier.

[0081] Fig. 10 shows another example of a scene 80 including a labelled and selected visual object 90 and several examples for selecting a proper modification of the corresponding audio source. For example, the above method may further comprise the steps of detecting at least one gesture 102 of the user on or above the display 30, wherein the marked visual object 88 may be selected in accordance with the gesture 102 and/or the audio signal may be selectively modified with respect to the sound from the audio source associated with the selected visual object 90 in accordance with the gesture 102. The gesture 102 may be a movement of the thumb relative to the forefinger, wherein an increase of the distance between the thumb and the forefinger may be representative for an increase of the value of the corresponding parameter for modifying the audio signal with respect to the selected visual object 90 and/or a decrease of the distance between the thumb and the forefinger may be representative for a decrease of the value of the corresponding parameter for modifying the audio signal with respect to the selected visual object 90. Further, a beamformer symbol 104, e.g. a triangle, may be laid over the selected visual object 90 and the user may increase or decrease an upper base of the triangle and as such the value of the corresponding parameter by the gesture 102. Alternatively or additionally the slider 100 may be provided on the display 30 within the scene 80, wherein the slider 100 may be used to increase or decrease the value of one of the parameters for modifying the audio signal.

[0082] Fig. 11 shows an example of a visual object-based audio signal modifier implementation. In particular, figure 11 shows an example of a visual object Obj_k acting as an audio source, wherein the visual object Obj_k may be taken from a lookup table. When such a visual object Obj_k is detected, the corresponding acoustical information is analysed and the most suitable sub-object $Obj_{k,1}$, $Obj_{k,2}$, $Obj_{k,3}$ with its set of weights w_a, ..., w_g may be chosen, e.g. depending on the corresponding acoustical context. The weight matrices may be added to the modifiers to consider the object-depended impact of each modifier as well as the acoustical context of the corresponding object Obj_k. So, each visual object Obj_k may have an individual set of weights w_a, ..., w_g, which depends also on the acoustical information of the visual object. In addition, each visual object Obj_k may have the same modification possibilities Mod_a, ..., Mod_g. Depending on the visual object Obj_k and, in case the acoustic environment is classified, the acoustic feature activity the relevant modification parameters will be chosen.

[0083] One possible realization could be the set of weights w_a, ..., w_g of the predefined modifiers Mod_a, ..., Mod_g as shown in figure 11. For each connection, there may be one of the weights w_a, ..., w_g, which depends on the detected visual object Obj_k. This means, for each visual object and/or object class in the database there is the set of weights w_a, ..., w_g. If the user wants to increase the sound from one of the visual objects Obj_k, there are modifiers, which may help for this purpose.

[0084] For example, the appropriate weights w_a, ..., w_g may get a rather high value close to 1, which means that the visual object-based modifier may have a high impact on this modifier for this specific visual object or object class. If the modifier is expected to not help for this purpose, the appropriate weights w_a, ..., w_g may get a rather low value close to 0, which means that the sound object-based modifier will have a low impact or no impact on this modifier for this specific visual object and/or object class. It is also possible to set the weights w_a, ..., w_g in the mid-range so that there is a moderate impact on this modifier for a specific visual object and/or object class. A level dependency or other properties of the acoustic environment may be added to consider that the impact of each modifier on the same visual object 88 may have a different strength depending on the sound properties of the visual object 88 and the corresponding audio source as well as the acoustic environment.

[0085] While the invention has been illustrated and described in detail in the drawings and foregoing description, such illustration and description are to be considered illustrative or exemplary and not restrictive; the invention is not limited to the disclosed embodiments. Other variations to the disclosed embodiments can be understood and effected by those skilled in the art and practicing the claimed invention, from a study of the drawings, the disclosure, and the appended claims. In the claims, the word "comprising" does not exclude other elements or steps, and the indefinite article "a" or "an" does not exclude a plurality. A single processor or controller or other unit may fulfil the functions of several items recited in the claims. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage. Any reference signs in the claims should not be construed as limiting the scope.

LIST OF REFERENCE SYMBOLS

[0086]

10	hearing system
12	hearing device
14	user device
15	part behind the ear
16	part in the ear
18	tube
20	sound input module
22	sound output module
24	input mean
30	display
32	graphical user interface

34	control element
36	camera
40	first processing unit
42	sound processing module
44	control module
46	sound source detector
48	first classifier
50	first memory
52	first transceiver
60	second processing unit
62	second memory
64	second transceiver
66	second classifier
68	visual object detector
70	Internet
72	server
80	scene
82	person
84	instrument
86	speaker
88	visual object
90	selected visual object
92	label
94	state indicator
96	input field
98	button
100	slider
102	gesture
104	beamformer symbol
S2-S20	steps two to twenty

Claims

1. A method for operating a hearing system (10), the hearing system (10) comprising a hearing device (12) configured to be worn at an ear of a user,

a user device (14) communicatively coupled to the hearing device (12) and comprising a camera (36) and a display (30),
the hearing device (12) comprising at least one sound input module (20) for generating an audio signal indicative of a sound detected in an environment of the hearing device, a first processing unit (40) for modifying the audio signal, and at least one sound output module (22) for outputting the modified audio signal,
the method comprising:

receiving image data from the camera (36), the image data being representative of a scene (80) in front of the camera (36);
receiving an audio signal from the at least one sound input module (20), the audio signal being representative of the acoustic environment of the hearing device (12) substantially at a time the image data have been captured, wherein the acoustic environment comprises at least one audio source and wherein the audio signal is at least in part representative of a sound from the audio source;
determining at least one visual object (88) as the audio source within the scene (80) from the image data and the audio signal.

2. The method of claim 1, the method further comprising the steps of

displaying the scene (80) on the display (30);
marking the determined visual object (88) within the scene (80);
receiving an input of the user to select the marked visual object (88);
selectively modifying the audio signal from the audio source associated with the selected visual object (90); and
outputting the modified audio signal to the user.

3. The method of one of claims 1 or 2, wherein the step of determining at least one visual object (88) as the audio source, within the scene (80) from the image data and the audio signal comprises:

determining at least one visual object (88), which is a potential audio source, within the scene (80) from the image data;
determining at least one audio source within the acoustic environment from the audio signal;
associating at least one determined visual object (88) with at least one determined audio source.

4. The method of claim 3, wherein

the step of determining the at least one visual object (88), which is the potential audio source, comprises determining a first spatial relationship between the camera (36) and the visual object (88);
the step of determining the at least one audio source within the acoustic environment comprises determining a second spatial relationship between the hearing device (12) and the audio source; and
the step of associating the at least one determined visual object (88) with the at least one determined audio source comprises

comparing the first spatial relationships of all determined visual objects (88) with the second spatial relationships of all determined sound sources, and
associating that visual object (88) of all determined visual objects (88) with that audio

source of all determined sound sources such that the corresponding first and second spatial relationship fulfil a predetermined requirement.

5. The method of one of the previous claims, further comprising:

classifying the acoustic environment from the received audio signal;
modifying the audio signal in accordance with the classification; and
determining the at least one audio source within the acoustic environment from the modified audio signal.

6. The method of one of the previous claims, further comprising:

determining at least one object class of the determined visual object (88) and
labelling the marked visual object (88) in the scene (80) in accordance with the determined object class.

7. The method of one of the previous claims, further comprising:
providing at least one input field (96) on the display (30) for the input of the user, with the input field (96) being representative for at least one modification of the audio signal with respect to the sound from the audio source assigned to the selected visual object (90), wherein the audio signal is selectively modified with respect to the sound from the audio source assigned to the selected visual object (90), if the user activates the input field (96).

8. The method of claim 7, wherein

the input field (96) is provided depending on the classification of the acoustic environment, with the input field (96) being representative for at least one modification in accordance with the classification of the acoustic environment; and/or
the input field (96) is provided depending on the object class of the determined visual object (88), with the input field (96) being representative for at least one modification in accordance with the object class of the determined visual object (88).

9. The method of one of the previous claims, wherein

the input is representative for the user wishing to increase the volume of the sound from the selected visual object (90) and/or to decrease the volume or effectuate a dampening of the sound from all other determined visual objects (88), or to decrease the volume or effectuate a dampening of the sound from the selected visual object (90) and/or to increase the volume of the sound from all other determined visual objects (88), and
the audio signal is selectively modified such that the volume of the audio source associated with the selected visual object (90) is increased or, respectively, decreased, or that the volume of the audio sources of all other determined visual objects (88) is decreased or, respectively, increased.

10. The method of one of the previous claims, further comprising:
monitoring the visual object (88) by the camera (36) and stopping to selectively modify the audio signal with respect to the sound from the audio source assigned to the monitored visual object (88) if the visual object (88) disappears from a field of view of the camera (36).

11. The method of one of the previous claims, further comprising:

detecting at least one gesture (102) of the user on or above the display (30); and
selecting the marked visual object (88) in accordance with the gesture (102); and/or
selectively modifying the audio signal with respect to the sound from the audio source associated with the selected visual object (90) in accordance with the gesture (102).

12. Hearing system (10), comprising:

a hearing device (12) configured to be worn at an ear of a user and comprising at least one sound input module (20) for generating an audio signal, a first processing unit (40) for modifying the audio signal, and at least one sound output module (24) for outputting the modified audio signal; and
a user device (14) communicatively coupled to the hearing device (12) and comprising a camera (36), a display (30), and a second processing unit (60),

wherein at least control unit is coupled to the hearing device (12) and the user device (14) and is configured for carrying out the method in accordance with one of the previous claims.

13. The hearing system (10) of claim 12, further comprising:
a remote server (72) communicatively coupled to the hearing device (12) and/or the user device (14) and being configured to carry out at least a part of the

method in accordance with one of claims 1 to 11.

14. The hearing system (10) of one of claims 12 or 13, wherein the control unit is implemented in the first processing unit (40), the second processing unit (60) or the remote server (72).

15. A computer program for operating a hearing system (10), which program, when being executed by a processing unit (40, 60), is adapted to carry out the steps of the method of one of claims 1 to 11.

16. A computer-readable medium, in which a computer program according to claim 15 is stored.

Fig. 1

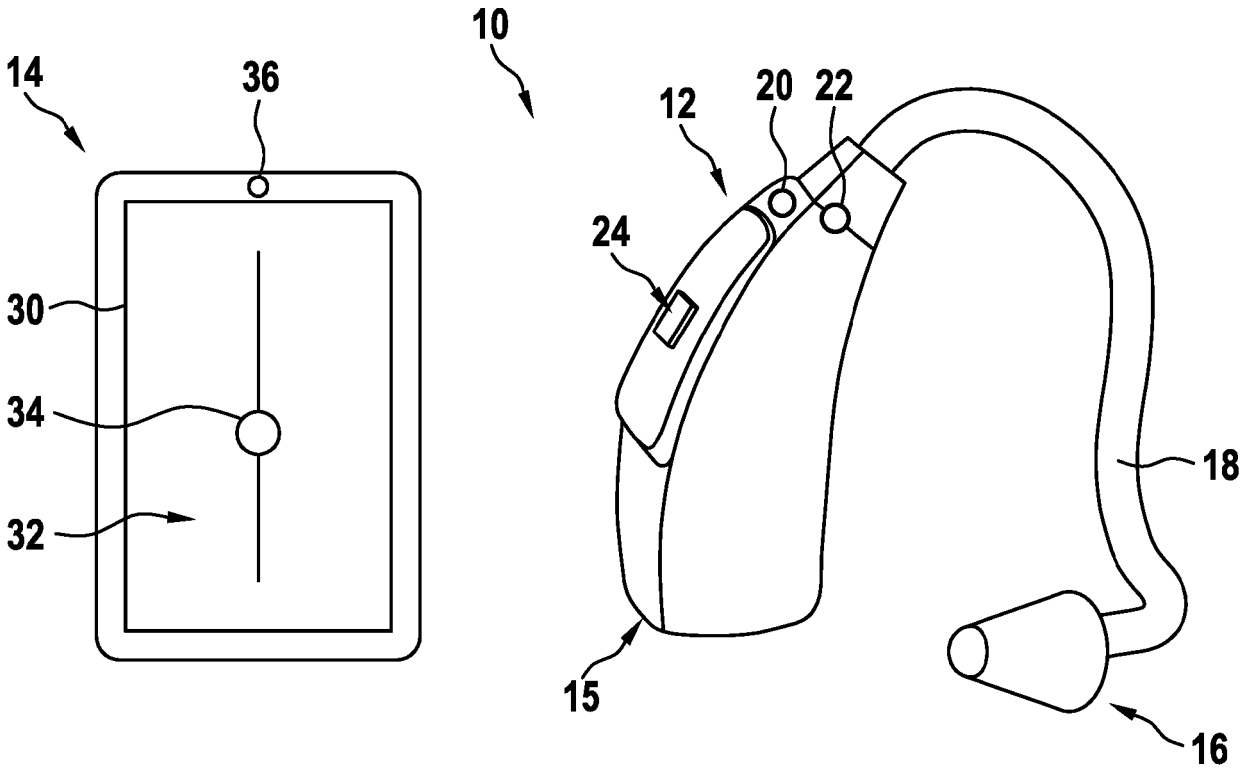


Fig. 2

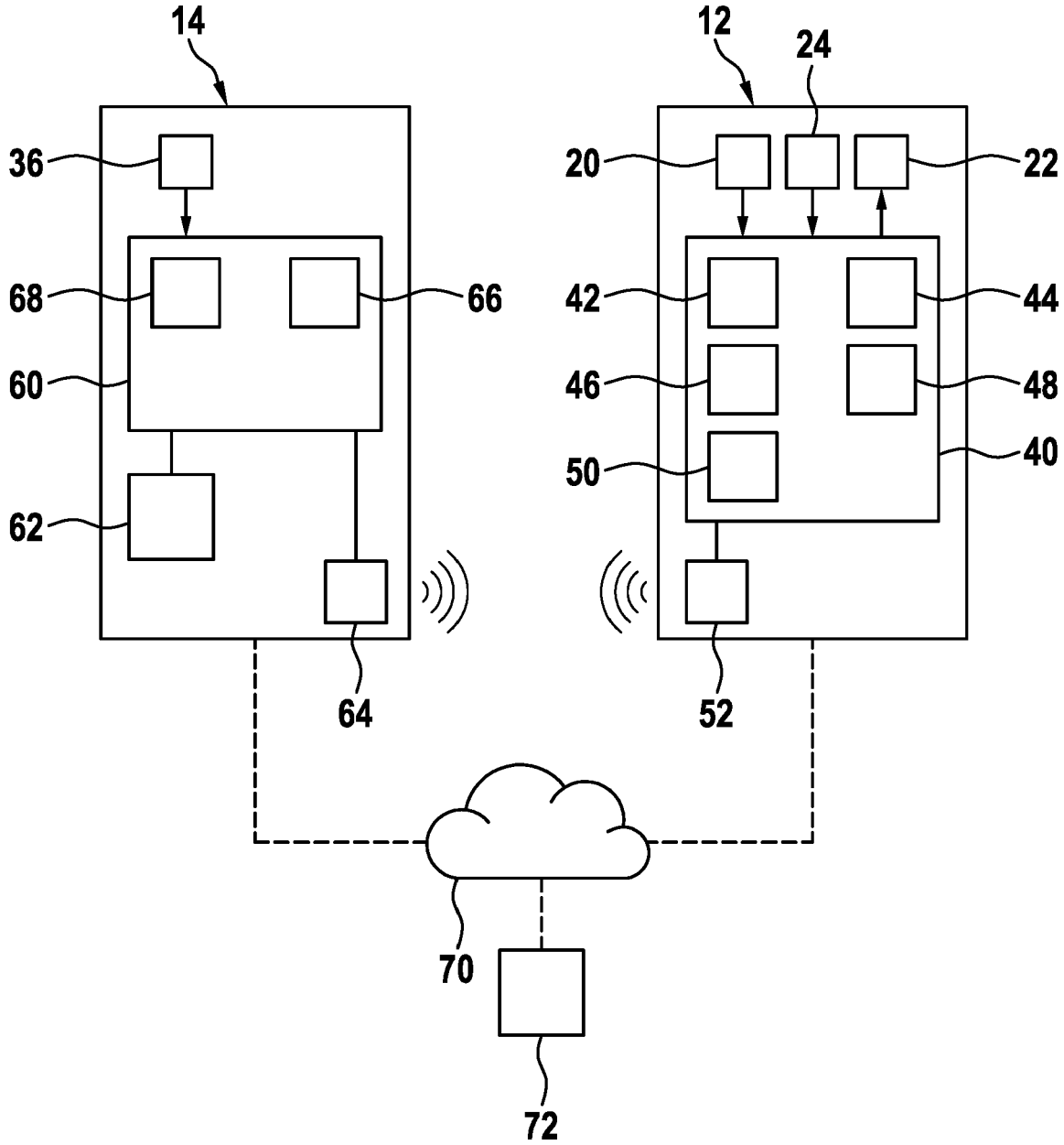


Fig. 3

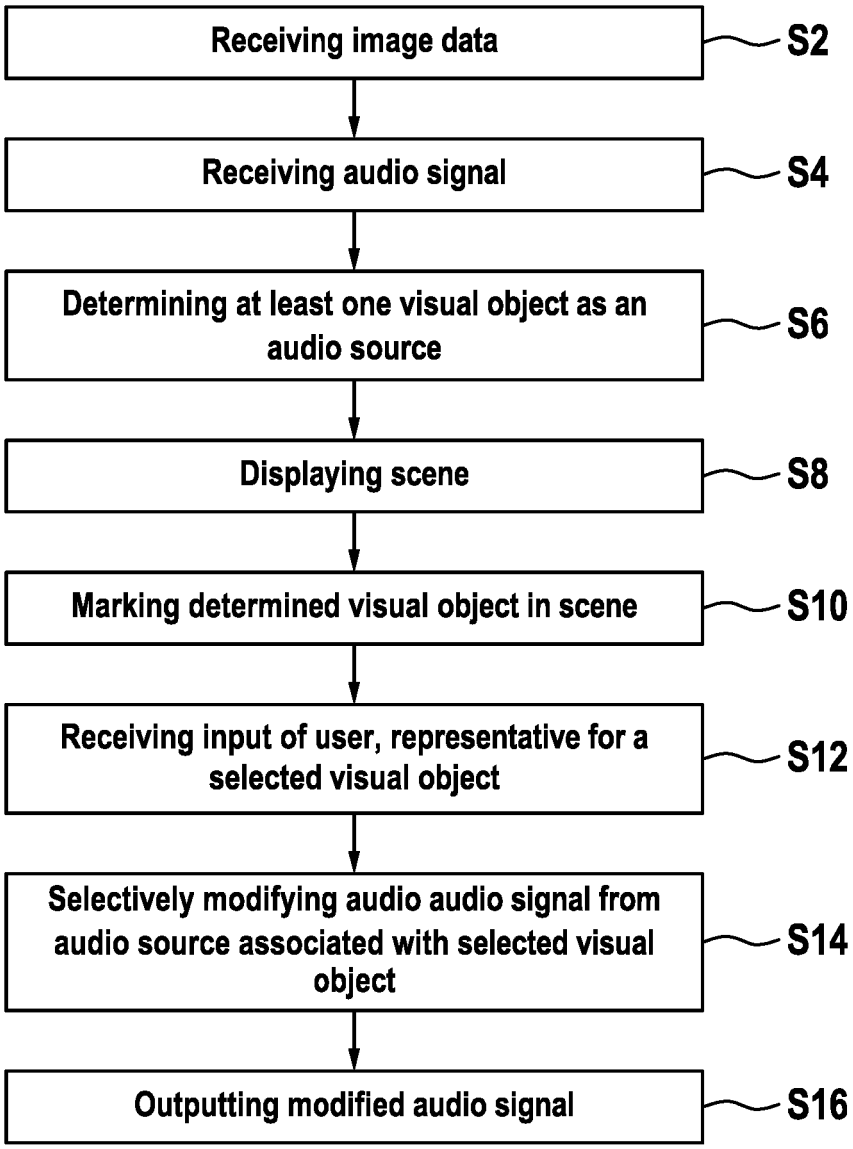

Receiving image data
S2
Receiving audio signal
S4
Determining at least one visual object as an audio source
S6
Displaying scene
S8
Marking determined visual object in scene
S10
Receiving input of user, representative for a selected visual object
S12
Selectively modifying audio audio signal from audio source associated with selected visual object
S14
Outputting modified audio signal
S16

Fig. 4

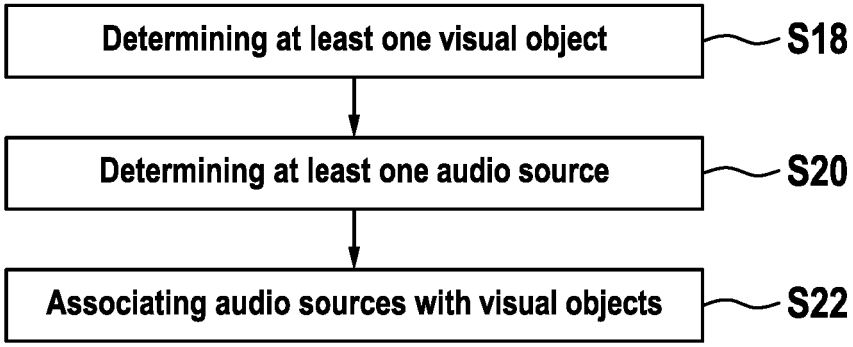

Determining at least one visual object
S18
Determining at least one audio source
S20
Associating audio sources with visual objects
S22

Fig. 5

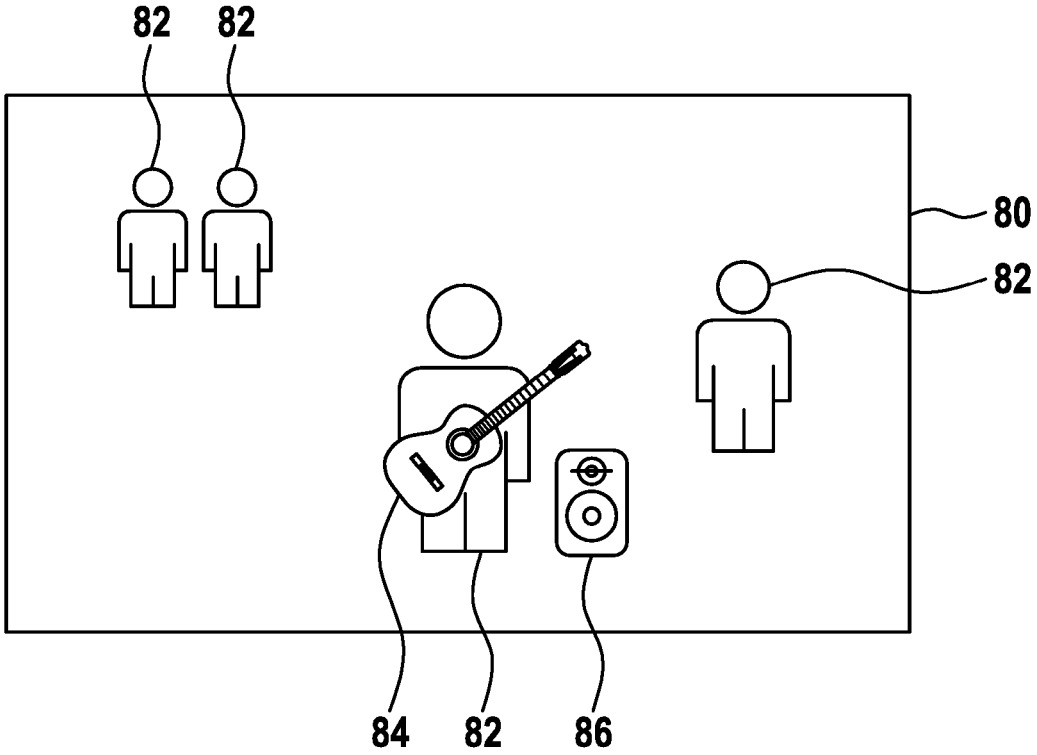


Fig. 6

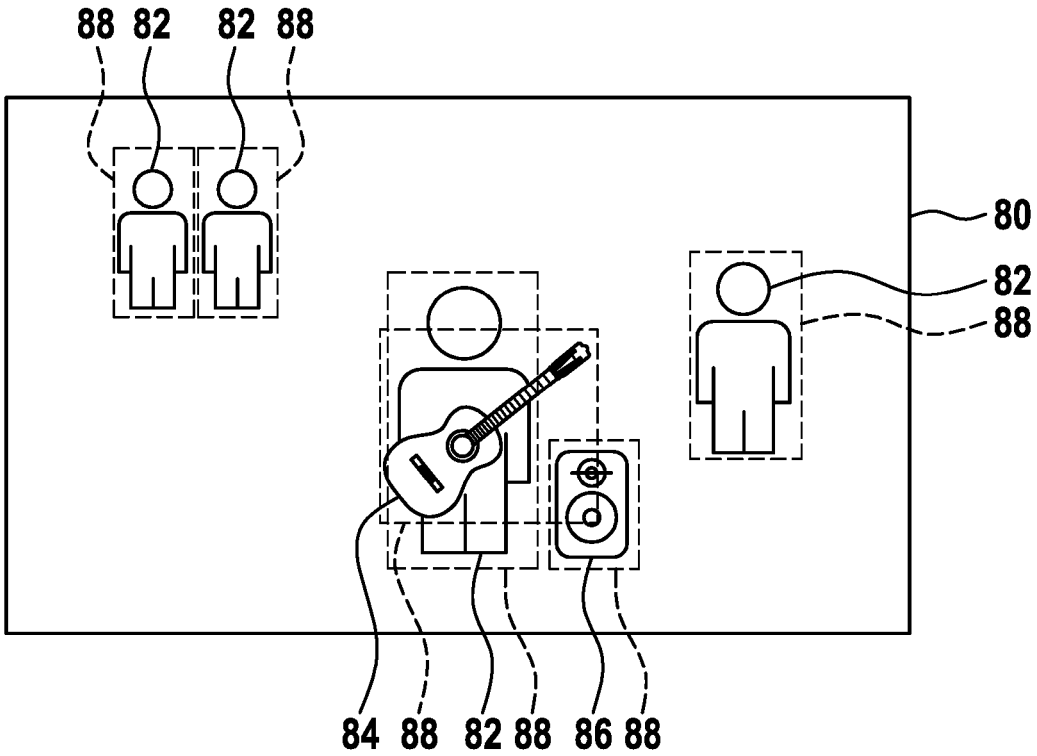


Fig. 7

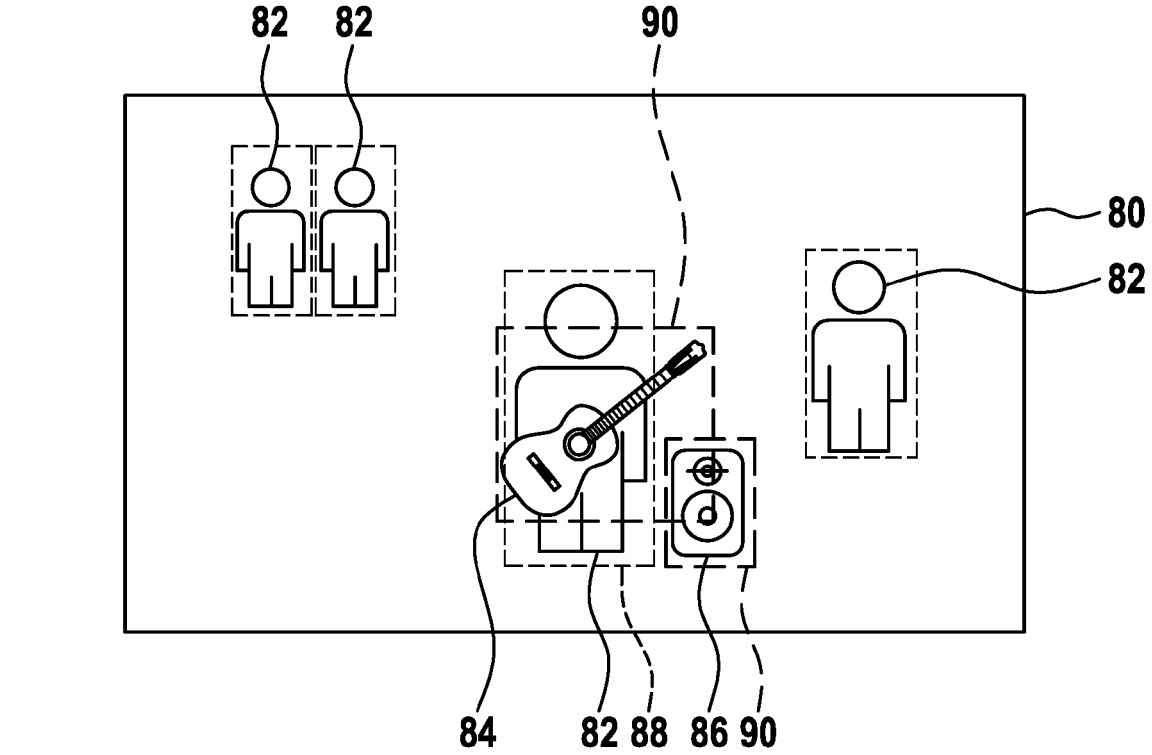

82
82
90
80
82
84
82
88
86
90

Fig. 8

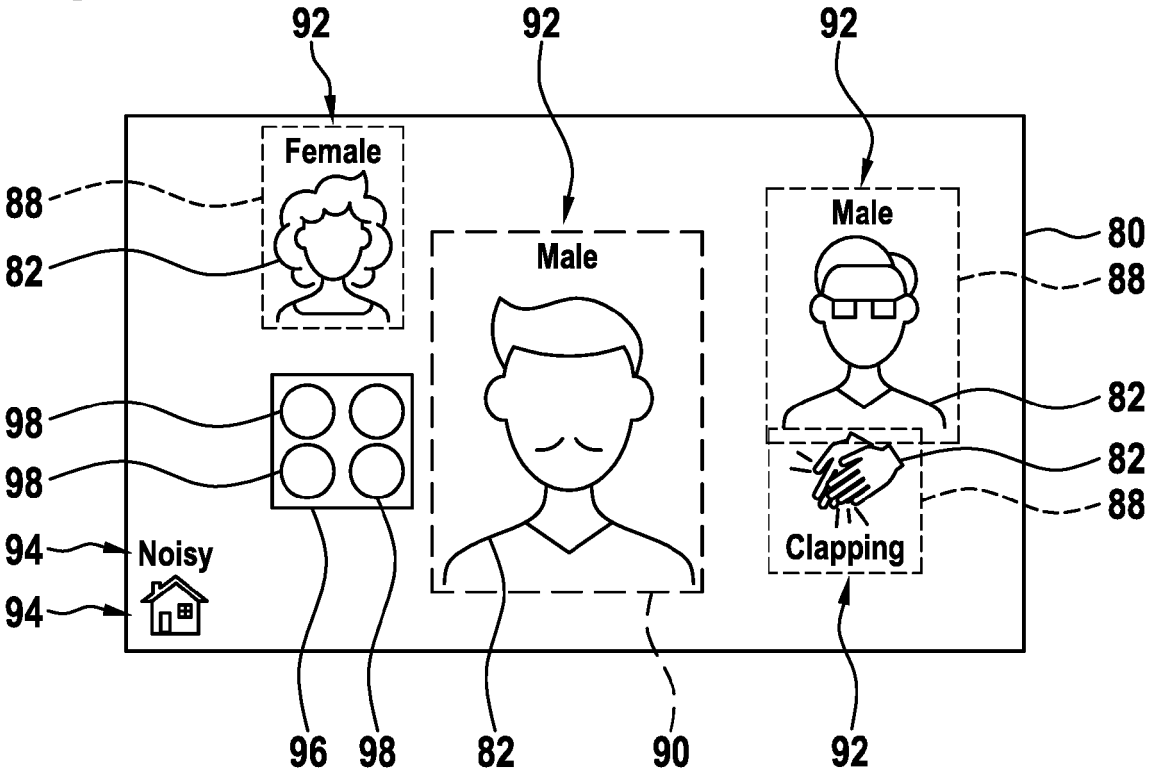

92
92
92
Female
Male
Male
88
82
80
88
98
98
82
82
94
Noisy
94
Clapping
88
96
98
82
90
92

Fig. 9

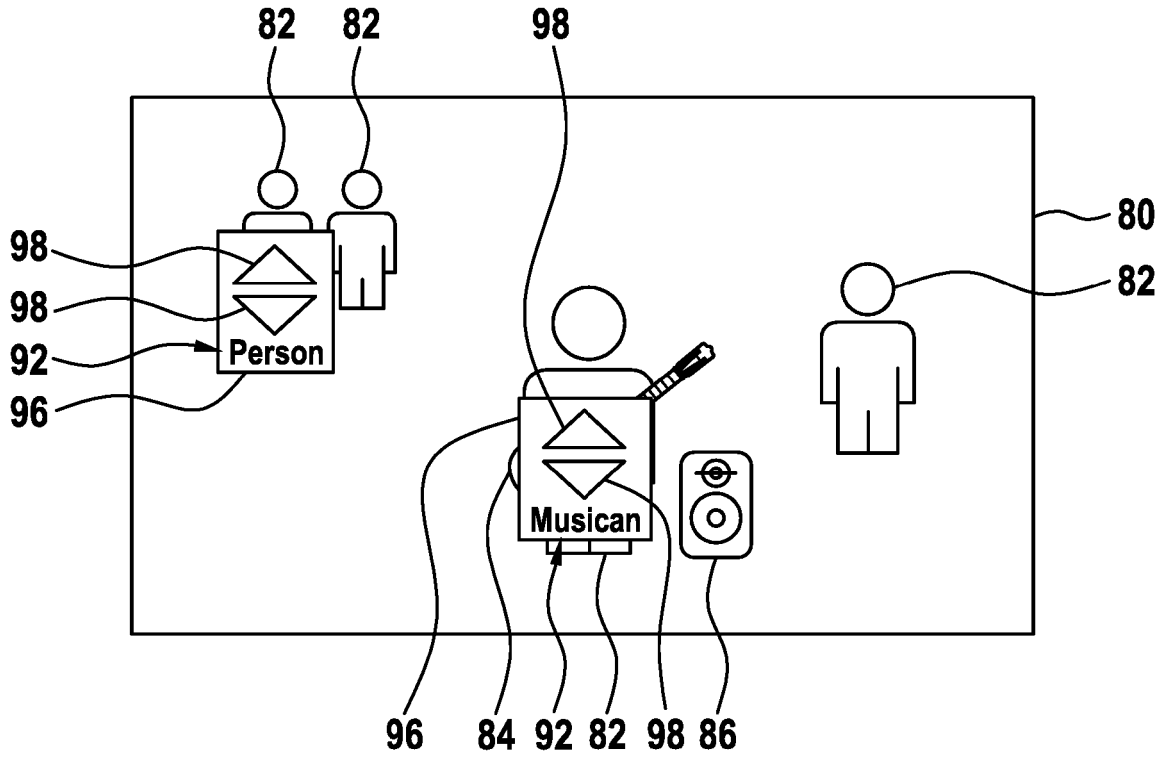

82
82
98
98
98
92
96
Person
Musican
80
82
96
84
92
82
98
86

Fig. 10

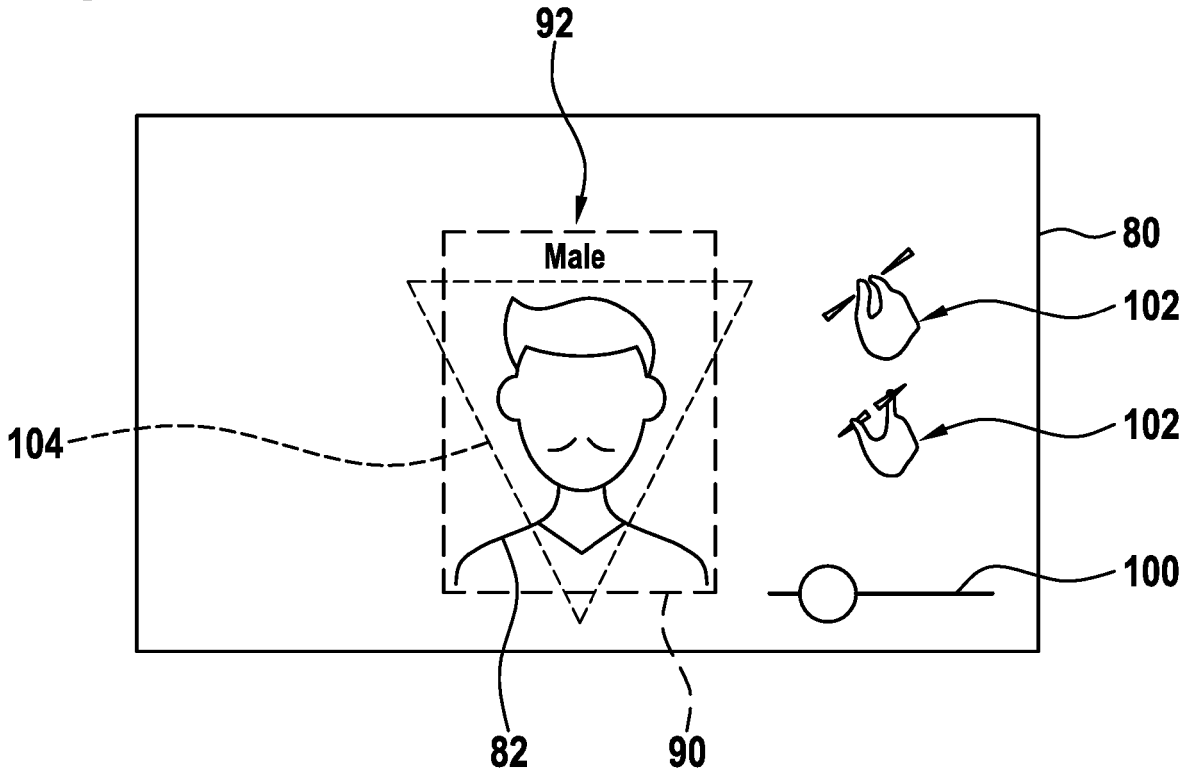
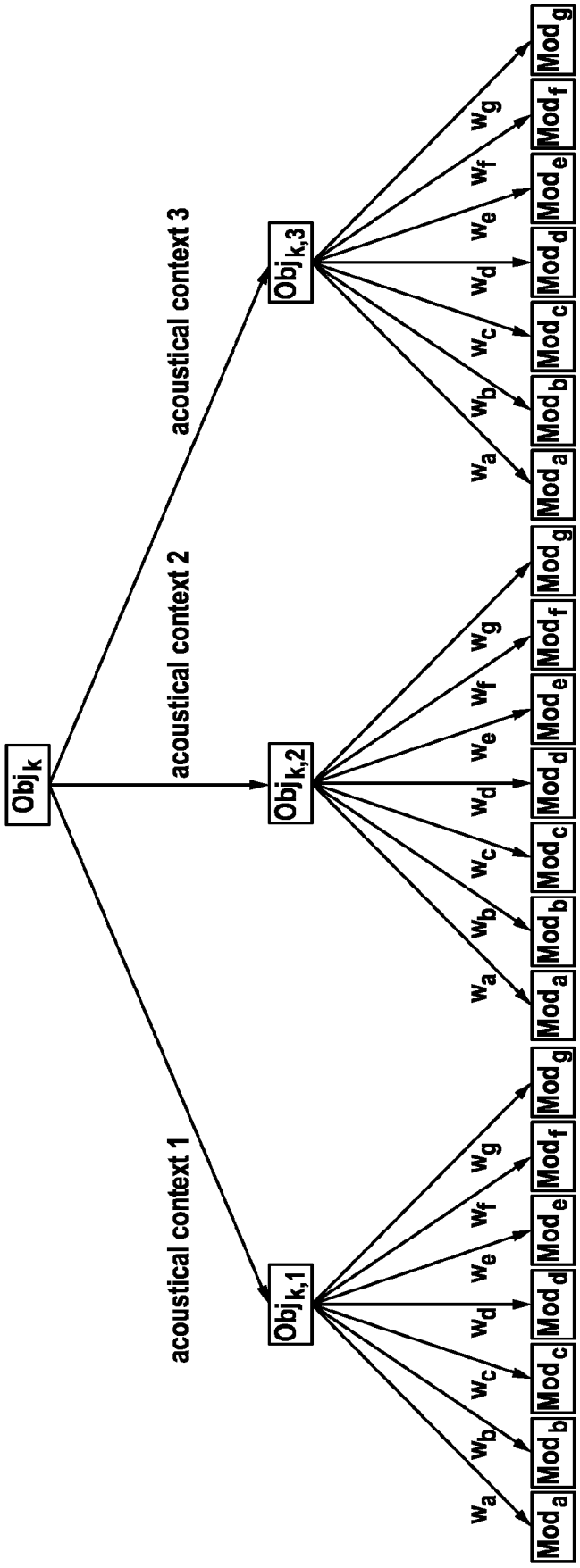

92
Male
80
102
102
104
100
82
90

Fig. 11


Obj_k
acoustical context 1
acoustical context 2
acoustical context 3
Obj_k,1
Obj_k,2
Obj_k,3
w_a
w_b
w_c
w_d
w_e
w_f
w_g
Mod_a
Mod_b
Mod_c
Mod_d
Mod_e
Mod_f
Mod_g

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 21 20 6983

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2018/088900 A1 (GLASER WILLIAM [US] ET AL) 29 March 2018 (2018-03-29) * paragraphs [0020] - [0025], [0029], [0030], [0034], [0035], [0044] - [0050], [0056], [0059] - [0064], [0073], [0074], [0078], [0087], [0088]; figures 1,2,5 *	1-3,9-16	INV. H04R25/00
X A	US 2016/277850 A1 (LI SCOTT WENTAO [US] ET AL) 22 September 2016 (2016-09-22) * paragraphs [0003], [0045], [0047], [0055] - [0059], [0064], [0076]; figures 2,6,7 *	1-3,6,7, 9,11-16 8	
X Y A	WO 2021/136962 A1 (ORCAM TECHNOLOGIES LTD [IL]) 8 July 2021 (2021-07-08) * paragraphs [0194], [0275], [0285], [0288], [0348], [0352], [0364], [0372], [0384] - [0389], [0393], [0394]; figures 38,39 * * paragraphs [0408], [0417], [0431], [0436] *	1,5,7,9, 11-16 10 8	TECHNICAL FIELDS SEARCHED (IPC) H04R G06K G06V G01S
X	US 2017/188173 A1 (RANIERI JURI [CH] ET AL) 29 June 2017 (2017-06-29) * paragraphs [0024], [0069] - [0074], [0086] - [0090], [0095] - [0098], [0112], [0114], [0118]; figures 1,2 *	1,3,4, 12-16	
Y A	WO 2021/038295 A1 (ORCAM TECHNOLOGIES LTD [IL]) 4 March 2021 (2021-03-04) * paragraphs [0004], [0102], [0184], [0189], [0197]; figures 26A, 26B *	10 1-9, 11-16	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	9 May 2022	Carrière, Olivier

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)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 21 20 6983

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

09-05-2022

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2018088900 A1	29-03-2018	US 2018088900 A1	29-03-2018
		US 2019354343 A1	21-11-2019
		US 2020192629 A1	18-06-2020
		US 2021405960 A1	30-12-2021
US 2016277850 A1	22-09-2016	NONE	
WO 2021136962 A1	08-07-2021	NONE	
US 2017188173 A1	29-06-2017	NONE	
WO 2021038295 A1	04-03-2021	NONE	

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82