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

(11) **EP 4 339 947 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
20.03.2024 Bulletin 2024/12

(21) Application number: **22196043.8**

(22) Date of filing: **16.09.2022**

(51) International Patent Classification (IPC):
G10L 25/27 *(2013.01)* *G10L 21/02* *(2013.01)*
G10L 25/30 *(2013.01)* *H04R 25/00* *(2006.01)*
H04R 5/04 *(2006.01)*

(52) Cooperative Patent Classification (CPC):
H04R 25/70; G10L 21/02; **G10L 25/27; H04R 5/04;**
H04R 25/50; G10L 25/30; H04R 2460/11

(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: **GN Audio A/S**
2750 Ballerup (DK)

(72) Inventor: **MOWLAEE, Pejman**
2750 Ballerup (DK)

(74) Representative: **Zacco Denmark A/S**
Arne Jacobsens Allé 15
2300 Copenhagen S (DK)

(54) METHOD FOR DETERMINING ONE OR MORE PERSONALIZED AUDIO PROCESSING PARAMETERS

(57) A method for determining one or more personalized audio processing parameters for an audio device is disclosed, and an audio device configured to carry out a corresponding method. The method comprises the steps of obtaining an input audio signal, obtaining one or more user parameters, determining an error function, determining a perceptual constraint, determining one or more optimized audio processing parameters by minimizing the error function constrained by the perceptual constrain.

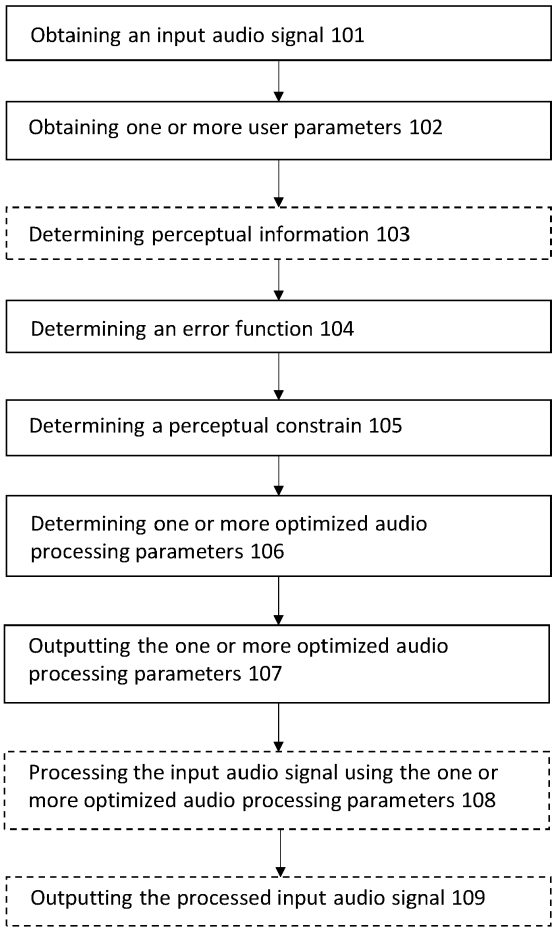

Fig. 3

EP 4 339 947 A1

Description

[0001] The present disclosure relates to a method for determining one or more personalized audio processing parameters, and a related audio device.

BACKGROUND

[0002] Most audio devices used for communication are provided with digital signal processing (DSP) capabilities allowing for the audio devices to modify obtained audio signals. The obtained audio signal may be audio signals obtained by a microphone of the audio device, or an audio signal received from a far-end device. The modification of an audio signal may comprise to enhance the quality of speech within the audio signal, for example, by removing noise or reducing echo in the audio signal.

[0003] Processing carried out by DSP on obtained audio signals is generally carried out by providing the best possible enhancement of the audio signal having in mind the average hearing profile of people. Alternatively, the DSP of an audio device may be tuned during a tuning process, where one or more audio engineers try to fine tune one or more parameters of the DSP. Both approaches are of a general nature where an one-size-fits-all approach is taken to DSP.

[0004] However, the hearing profile and listening preferences of end users of audio device may vary drastically, thus, the above-mentioned approaches suffer from being too general in their approach to signal processing, as none of these method delivers a solution where the DSP is optimized with the end target user in mind. Consequently, the general approaches to DSP tuning or parameter setting may give acceptable results to some, while for others it may lead to unacceptable results.

[0005] To overcome this problem systems and methods for modifying an audio signal using custom psychoacoustic models have been proposed. WO2020016440 discloses systems and methods for modifying an audio signal using custom psychoacoustic models. A user's hearing profile is first obtained. Subsequently, a multiband dynamic processor is parameterized to optimize the user's perceptually relevant information. The method for calculating the user's perceptually relevant information comprises first processing audio signal samples using the parameterized multiband dynamic processor and then transforming samples of the processed audio signals into the frequency domain. Next, masking and hearing thresholds are obtained from the user's hearing profile and applied to the transformed audio sample, wherein the user's perceived data is calculated. Once perceptually relevant information is optimized, the resulting parameters are transferred to a multiband dynamic processor and an output audio signal is processed.

[0006] However, the presented method still suffers from being slow and cumbersome, preventing it from being used in real-time and to adapt on the fly to obtained audio signals. Furthermore, the method is limited in the amount of customization being providable to an end user.

SUMMARY

[0007] Accordingly, there is a need for audio devices and methods with improved capabilities for determining one or more personalized audio processing parameters.

[0008] According to a first aspect of the disclosure there is provided a method for determining one or more personalized audio processing parameters for an audio device comprising an interface, a speaker, and one or more microphones, the method comprising:

obtaining an input audio signal,

obtaining one or more user parameters indicative of one or more characteristics of a target user,

determining an error function, wherein the error function is a function of a target audio signal, the input audio signal and the one or more audio processing parameters,

determining a perceptual constraint based on the one or more user parameters,

determining one or more optimized audio processing parameters by minimizing the error function constrained by the perceptual constraint, and wherein the one or more audio processing parameters are the optimization variables, and

outputting the one or more optimized audio processing parameters.

[0009] Consequently, an improved method for determining one or more personalized audio processing parameters is

provided. The provided method allows for a fast and precise determination of one or more optimized audio processing parameters based on the target user. The provided method is suitable for making on-the-fly determinations of optimized audio processing parameters, i.e., optimized audio processing parameter may be determined during real-time processing in online conferences or meetings. Furthermore, the provided method allows for a high degree of customization allowing for the optimized parameters to be optimized both towards the end users listening preferences, but also or alternatively to the hearing profile of the end user.

[0010] In the context of the present disclosure, the term personalized, or personalizing is to be construed as something being done to cater to the user using the audio device, e.g., a user wearing a headset where audio being played through the headset is processed based on one or more characteristics of the user wearing the headset. In the present disclosure personalization is achieved by determining one or more optimized audio processing parameters to cater to the hearing profile or listening preferences of an end target user.

[0011] The one or more audio processing parameter may be any parameters related to processing of the audio signal. The one or more audio processing parameters may be one or more audio filter quotients. The one or more audio processing parameters may be one or more gain values. The one or more audio processing parameters may be one or more audio masks. The one or more audio processing parameters may be one or more weights, e.g., weights of a neural network or weights of an audio filter. The one or more audio processing parameters may be one or more mixing values. The one or more audio processing parameters may be one or more nodes within a neural network. The one or more audio processing parameters may be neural network architecture. The one or more audio processing parameters may be a selection of a processing pipeline.

[0012] The input audio signal may be obtained in a plurality of manners. The input audio signal may be received from a far-end station, such as another audio device or a server device. The input audio signal may be obtained by retrieving the input audio signal from a local storage on the audio device. The local storage may be a memory of the audio device. The input audio signal may be part of an online conference between a far-end device, and a near-end device. The input audio signal may be a test signal stored on the audio device. The input audio signal may be obtained by one or more microphone of the audio device. The input microphone signal may be a media signal. A media signal may be a signal representative of a song, audio of a movie or an audio book. The input microphone signal may be voice signal recorded during a phone call or another communication session between two or more parties. The input microphone signal may be a signal obtained in real-time, e.g., the input microphone signal being part of an on-going online conference. The input microphone signal may be part of a larger dataset of input microphone signals. The input audio signal may be a time domain signal or a frequency domain signal. The input audio signal may be obtained via the processor of the audio device.

[0013] In some embodiments a plurality of input audio signal may be obtained and used in determining the one or more optimized audio processing parameters. Preferably, one set of optimized audio processing parameters is determined in real-time with obtaining the input audio signal. In an embodiment a first input audio signal is obtained at a first time, and one or more first optimized audio processing parameters are determined based on the first input audio signal, and a second input audio signal is obtained at a second time, and one or more second optimized audio processing parameters are determined based on the second input audio signal.

[0014] The one or more user parameters may be obtained by receiving one or more inputs from a target user. The one or more user parameters may be obtained by retrieving the one or more user parameters from a local storage on the audio device, such as a flash drive. The one or more user parameters may be obtained by retrieving the one or more user parameters from an online profile of the user, e.g., a user profile stored on a cloud. The one or more user parameters may be obtained by a hearing test carried out on the audio device. The one or more user parameters may be obtained from an external device communicatively connected to the audio device. For example, a hearing test may have been carried out on an external device, the result of the hearing test may then be transmitted to the audio device or uploaded to a server to allow for the audio device to retrieve the one or more user parameters. The one or more user parameters are representative of one or more characteristics of the target user. The one or more user parameters may be obtained via the processor of the audio device. The one or more user parameters may be updated with one or more new user parameters, e.g., if the listening preferences or the hearing profile of the user changes, the one or more user parameters may be updated accordingly. The one or more user parameters may be obtained in real-time during when using the audio device, e.g., during an online conference. The one or more user parameters may be obtained prior to a session where the user needs to use the audio device. The one or more user parameters may be obtained when starting the audio device or when starting using the audio device, e.g., for an online conference.

[0015] The one or more characteristics of the target user may be related to a user's audio preferences, e.g., if the user prefer a high gain on bass or treble. The one or more characteristics of the user may be related to the user themselves, e.g., a hearing loss, physiological data, a preferred wear style of the audio device, or other. The one or more characteristics of the user may be one or more user preferences. The one or more characteristics of the target user may be related to a user's sensitivity to different types of audio degradation, e.g., whether the user is highly sensitive to echo, or noise. The one or more characteristics of the target user may be one or more listening preferences of the target user.

[0016] The target user in the present disclosure is to be understood as the user for which the one or more optimized audio processing parameters are determined. The target user may be an intended recipient of the input audio signal. The target user may be an intended recipient of the processed input audio signal, wherein the input audio signal has been processed with the one or more optimized audio processing parameters. The target user may be a user of the audio device, e.g., if the audio device receives an input audio signal from a different device, the audio device may determine the one or more optimized audio processing parameters and process the input audio signal using the one or more optimized audio processing parameters, thereby providing an audio signal which have been processed in a personalized manner to suit the listening preferences or hearing profile of the user of the audio device. The target user may be a user of another device communicatively connected to the audio device such as another audio device, an electronic device or a server device, e.g., if the audio device is used for an online meeting the audio device may obtain the input audio signal via a microphone of the audio device, determine the one or more optimized processing parameters for one or more other participants in the meeting, process the obtained input audio signal, and transmit it to the one or more other participants in the meeting.

[0017] The error function may be any error function giving a measure of a quality of an estimator. The error function may be determined based on a Bayes estimator. The error function may be a linear function, e.g., the error function can be expressed as a mean square error or a least square error. The error function may be a non-linear function, e.g., a logarithm function (expander), exponent function (compressor), or a homomorphic transformation.

[0018] The target audio signal is in the present disclosure the parameter being estimated in the error function. The target audio signal may be a clean audio signal, i.e., the input audio signal without undesired distortions. The target audio signal may be dependent on the target user, i.e., the target audio signal may be dependent on the listening preferences and/or the hearing profile of the target user. The target audio signal may be defined as the input signal without undesired audio degradation.

[0019] The input audio signal and the one or more audio processing parameters may be combined to denote an estimator in the error function.

[0020] The error function may be determined by the processor of the audio device.

[0021] The perceptual constraint may be any weight and/or threshold based on the one or more user parameters. The perceptual constraint may be a set of weights associated with different audio distortion types, determining which type of audio distortion to penalize harder and which to penalize softer. The perceptual constraint may reflect preferences of the target user. The perceptual constraint may reflect a hearing profile, e.g., an audiogram, of the target user. The perceptual constraint may reflect a masking threshold of the user. The perceptual constraint may be a threshold for certain frequency ranges, such as bass, mid and/or treble. The perceptual constraint may be weight determining on which frequency ranges to put emphasis on, e.g., treble over bass. The perceptual constraint may reflect a user's sensitivity to different types of audio degradation. In general, the perceptual constraint may be viewed as any constrain constraining the solution space of the optimized audio processing parameters to fit with the user's listening preferences and/or sensitivity to different types of audio distortions. The perceptual constrain may be a weight shaping the input audio signal. The perceptual constraint may be based on a listening preference, and/or a hearing profile of the target user.

[0022] The perceptual constraint may be determined based on a hearing test performed on the user. The perceptual constraint may be determined based on the audible range and/or audible levels of the user. The perceptual constraint may be determined by a hearing test testing the user's preferences, e.g., if the user prefer bass or treble or if the user is especially sensitive to echo annoyance. The perceptual constrain may be determined based on the one or more user parameters. The perceptual constrain may be determined based on a questionnaire filled out by the target user on the listening preferences of the target user. As an example, given the masking threshold and the detection threshold of an end-user towards different audio distortions, such as echo annoyance or noise annoyance, a weighting matrix may be determined composed of frequency responses to shape the audio distortions to be masked, such a weighting matrix may constitute a perceptual constraint. The perceptual constrain may be determined by the processor of the audio device.

[0023] The one or more optimized audio processing parameters may be any audio processing parameters which have been selected to minimize the error function. In the context of the present disclosure the one or more optimized audio processing parameters should not be construed as a single set of one or more optimal audio processing parameters, e.g., solving the optimization problem of minimizing the error function may be done in plethora of different numerical manners, alternatively, different assumptions may be made to derive a closed-form solution. Each solution for minimizing the error function may each provide different optimized audio processing parameters; hence, the optimized audio processing parameters should not necessarily be viewed as a single set of optimal audio processing parameters but may instead be construed as a set of optimized audio processing parameters from plurality of sets of optimized audio processing parameters. For example, the error function when may comprise several different local minimum values for different audio processing parameters, and the optimized audio processing parameters may not necessarily lead to the global minimum value for the error function. The one or more optimized audio processing parameters may be determined by the processor of the audio device.

[0024] The processor of the audio device may output the one or more optimized audio processing parameters. Out-

putting the one or more optimized audio processing parameters may comprise replacing audio processing parameters of the audio device with the optimized audio processing parameters. Outputting the one or more optimized audio processing parameters may comprise transmitting the one or more optimized audio processing to a server device, another audio device, or an electronic device. Outputting the one or more optimized audio processing parameters may comprise replacing audio processing parameters of an audio filter of the audio device with the optimized audio processing parameters. Outputting the one or more optimized audio processing parameters may comprise backpropagating the optimized audio processing parameters through a neural network.

[0025] In one or more example audio devices, the interface comprises a wireless transceiver, also denoted as a radio transceiver, and an antenna for wireless transmission and reception of an audio signal, such as for wireless transmission of an output signal and/or wireless reception of a wireless input signal. The audio device may be configured for wireless communication with one or more electronic devices, such as another audio device, a smartphone, a tablet, a computer and/or a smart watch. The audio device optionally comprises an antenna for converting one or more wireless input audio signals to antenna output signal(s). The audio device may be configured for wireless communications via a wireless communication system, such as short-range wireless communications systems, such as Wi-Fi, Bluetooth, Zigbee, IEEE 802.11, IEEE 802.15, infrared and/or the like. The audio device may be configured for wireless communications via a wireless communication system, such as a 3GPP system, such as a 3GPP system supporting one or more of: New Radio, NR, Narrow-band IoT, NB-IoT, and Long Term Evolution - enhanced Machine Type Communication, LTE-M, millimeter-wave communications, such as millimeter-wave communications in licensed bands, such as device-to-device millimeter-wave communications in licensed bands. In one or more example audio devices the interface of the audio device comprises one or more of: a Bluetooth interface, Bluetooth low energy interface, and a magnetic induction interface. For example, the interface of the audio device may comprise a Bluetooth antenna and/or a magnetic interference antenna. In one or more example audio devices, the interface may comprise a connector for wired communication, via a connector, such as by using an electrical cable. The connector may connect one or more microphones to the audio device. The connector may connect the audio device to an electronic device, e.g., for wired connection. The one or more interfaces can be or comprise wireless interfaces, such as transmitters and/or receivers, and/or wired interfaces, such as connectors for physical coupling

[0026] In an embodiment, the audio device is configured to be worn by a user. The audio device may be arranged at the user's ear, on the user's ear, over the user's ear, in the user's ear, in the user's ear canal, behind the user's ear and/or in the user's concha, i.e., the audio device is configured to be worn in, on, over and/or at the user's ear. The user may wear two audio devices, one audio device at each ear. The two audio devices may be connected, such as wirelessly connected and/or connected by wires, such as a binaural hearing aid system.

[0027] The audio device may be a hearable such as a headset, headphone, earphone, earbud, hearing aid, a personal sound amplification product (PSAP), an over-the-counter (OTC) audio device, a hearing protection device, an one-size-fits-all audio device, a custom audio device or another head-wearable audio device. The audio device may be a speakerphone or a soundbar. Audio devices can include both prescription devices and non-prescription devices. The audio device may be a smart device, such as a smart phone.

[0028] The audio device may be embodied in various housing styles or form factors.

[0029] Some of these form factors are earbuds, on the ear headphones or over the ear headphones. The person skilled in the art is aware of different kinds of audio devices and of different options for arranging the audio device in, on, over and/or at the ear of the audio device wearer. The audio device (or pair of audio devices) may be custom fitted, standard fitted, open fitted and/or occlusive fitted.

[0030] In an embodiment, the audio device may comprise one or more input transducers. The one or more input transducers may comprise one or more microphones. The one or more input transducers may comprise one or more vibration sensors configured for detecting bone vibration. The one or more input transducer(s) may be configured for converting an acoustic signal into a first electric input signal. The first electric input signal may be an analogue signal. The first electric input signal may be a digital signal. The one or more input transducer(s) may be coupled to one or more analogue-to-digital converter(s) configured for converting the analogue first input signal into a digital first input signal.

[0031] In an embodiment, the audio device may comprise one or more antenna(s) configured for wireless communication. The one or more antenna(s) may comprise an electric antenna. The electric antenna may be configured for wireless communication at a first frequency. The first frequency may be above 800 MHz, preferably a wavelength between 900 MHz and 6 GHz. The first frequency may be 902 MHz to 928 MHz. The first frequency may be 2.4 to 2.5 GHz. The first frequency may be 5.725 GHz to 5.875 GHz. The one or more antenna(s) may comprise a magnetic antenna. The magnetic antenna may comprise a magnetic core. The magnetic antenna may comprise a coil. The coil may be coiled around the magnetic core. The magnetic antenna may be configured for wireless communication at a second frequency. The second frequency may be below 100 MHz. The second frequency may be between 9 MHz and 15 MHz.

[0032] In an embodiment, the audio device may comprise one or more wireless communication unit(s). The one or more wireless communication unit(s) may comprise one or more wireless receiver(s), one or more wireless transmitter(s), one or more transmitter-receiver pair(s) and/or one or more transceiver(s). At least one of the one or more wireless

communication unit(s) may be coupled to the one or more antenna(s). The wireless communication unit may be configured for converting a wireless signal received by at least one of the one or more antenna(s) into a second electric input signal. The audio device may be configured for wired/wireless audio communication, e.g., enabling the user to listen to media, such as music or radio and/or enabling the user to perform phone calls.

[0033] In an embodiment, the wireless signal may originate from one or more external source(s) and/or external devices, such as spouse microphone device(s), wireless audio transmitter(s), smart computer(s) and/or distributed microphone array(s) associated with a wireless transmitter. The wireless input signal(s) may origin from another audio device, e.g., as part of a binaural hearing system and/or from one or more accessory device(s), such as a smartphone and/or a smart watch.

[0034] In an embodiment, the audio device may include a processor. The processor may be configured for processing the first and/or second electric input signal(s). The processing may comprise compensating for a hearing loss of the user, i.e., apply frequency dependent gain to input signals in accordance with the user's frequency dependent hearing impairment. The processing may comprise performing feedback cancelation, echo cancellation, beamforming, tinnitus reduction/masking, noise reduction, noise cancellation, speech recognition, bass adjustment, treble adjustment and/or processing of user input. The processor may be a processor, an integrated circuit, an application, functional module, etc. The processor may be implemented in a signal-processing chip or a printed circuit board (PCB). The processor may be configured to provide a first electric output signal based on the processing of the first and/or second electric input signal(s). The processor may be configured to provide a second electric output signal. The second electric output signal may be based on the processing of the first and/or second electric input signal(s).

[0035] In an embodiment, the audio device may comprise an output transducer. The output transducer may be coupled to the processor. The output transducer may be a loudspeaker. The output transducer may be configured for converting the first electric output signal into an acoustic output signal. The output transducer may be coupled to the processor via the magnetic antenna.

[0036] In an embodiment, the wireless communication unit may be configured for converting the second electric output signal into a wireless output signal. The wireless output signal may comprise synchronization data. The wireless communication unit may be configured for transmitting the wireless output signal via at least one of the one or more antennas.

[0037] In an embodiment, the audio device may comprise a digital-to-analogue converter configured to convert the first electric output signal, the second electric output signal and/or the wireless output signal into an analogue signal.

[0038] In an embodiment, the audio device may comprise a vent. A vent is a physical passageway such as a canal or tube primarily placed to offer pressure equalization across a housing placed in the ear such as an ITE audio device, an ITE unit of a BTE audio device, a CIC audio device, a RIE audio device, a RIC audio device, a MaRIE audio device or a dome tip/earmold. The vent may be a pressure vent with a small cross section area, which is preferably acoustically sealed. The vent may be an acoustic vent configured for occlusion cancellation. The vent may be an active vent enabling opening or closing of the vent during use of the audio device. The active vent may comprise a valve.

[0039] In an embodiment, the audio device may comprise a power source. The power source may comprise a battery providing a first voltage. The battery may be a rechargeable battery. The battery may be a replaceable battery. The power source may comprise a power management unit. The power management unit may be configured to convert the first voltage into a second voltage. The power source may comprise a charging coil. The charging coil may be provided by the magnetic antenna.

[0040] In an embodiment, the audio device may comprise a memory, including volatile and non-volatile forms of memory.

[0041] The audio device may be configured for audio communication, e.g., enabling the user to listen to media, such as music or radio, and/or enabling the user to perform phone calls.

[0042] The audio device may comprise one or more antennas for radio frequency communication. The one or more antennas may be configured for operation in ISM frequency band. One of the one or more antennas may be an electric antenna. One or the one or more antennas may be a magnetic induction coil antenna. Magnetic induction, or near-field magnetic induction (NFMI), typically provides communication, including transmission of voice, audio, and data, in a range of frequencies between 2 MHz and 15 MHz. At these frequencies, the electromagnetic radiation propagates through and around the human head and body without significant losses in the tissue.

[0043] The magnetic induction coil may be configured to operate at a frequency below 100 MHz, such as at below 30 MHz, such as below 15 MHz, during use. The magnetic induction coil may be configured to operate at a frequency range between 1 MHz and 100 MHz, such as between 1 MHz and 15 MHz, such as between 1MHz and 30 MHz, such as between 5 MHz and 30 MHz, such as between 5 MHz and 15 MHz, such as between 10 MHz and 11 MHz, such as between 10.2 MHz and 11 MHz. The frequency may further include a range from 2 MHz to 30 MHz, such as from 2 MHz to 10 MHz, such as from 2 MHz to 10 MHz, such as from 5 MHz to 10 MHz, such as from 5 MHz to 7 MHz.

[0044] The electric antenna may be configured for operation at a frequency of at least 400 MHz, such as of at least 800 MHz, such as of at least 1 GHz, such as at a frequency between 1.5 GHz and 6 GHz, such as at a frequency between 1.5 GHz and 3 GHz such as at a frequency of 2.4 GHz. The antenna may be optimized for operation at a

frequency of between 400 MHz and 6 GHz, such as between 400 MHz and 1 GHz, between 800 MHz and 1 GHz, between 800 MHz and 6 GHz, between 800 MHz and 3 GHz, etc. Thus, the electric antenna may be configured for operation in ISM frequency band. The electric antenna may be any antenna capable of operating at these frequencies, and the electric antenna may be a resonant antenna, such as monopole antenna, such as a dipole antenna, etc. The resonant antenna may have a length of $\lambda/4 \pm 10\%$ or any multiple thereof, A being the wavelength corresponding to the emitted electromagnetic field.

[0045] In an embodiment the one or more user parameters comprises physiological information regarding the target user, such as gender and/or age.

[0046] Several studies have shown that hearing loss is well correlated with physiological parameters, such as age and gender. Thus, by obtaining relatively simple information regarding a target user the perceptual constrain may be customized based on the target user, thus leading to the derived optimized one or more audio processing parameters being customized to the target user. For example, based on the physiological information an estimation of the user's hearing profile may be made, which in turn may be used for determining the audible range and levels for the user and/or PRI, these values may be incorporated into the perceptual constrain, e.g., as threshold values. Physiological information regarding the user may be obtained by asking the user to input the information via a user interface, such as a user interface of a smart device or other electronic device communicatively connected to the audio device. Physiological information regarding the user may be obtained by asking the user to input the information via a user interface, the audio device may comprise the user interface. The physiological information regarding the user may comprise demographic information.

[0047] In an embodiment the one or more user parameters comprises a result of a hearing test carried out on the target user.

[0048] The result of the hearing test may be an audiogram. The result of the hearing test may be one or more listening preferences of the target user. The result of the hearing test may be one or more sensitivities the user has towards different audio distortion types, such as noise attenuation, speech distortion, dereverberation artifact, echo annoyance tolerance etc.. The result of the hearing test may be one or more listening preferences the user has, e.g., if the user prefer listening to audio which is bass boosted. The hearing test may be a professionally administrated hearing test carried out by an audiologist, such a test is normally carried out to identify hearing loss, audible ranges, and audible levels. The hearing test may a self-fitting hearing test carried out by the audio device itself. The hearing test may be a subjective listening test carried out by the audio device, e.g., where the audio device plays different audio clips to the user and the user rates the audio clips by giving an input to the interface of the audio device, the input may be in the form of a button press or a voice command. The hearing test may be a hearing test carried out by another audio device, and which is subsequently transmitted or otherwise transferred to the user's audio device.

[0049] In an embodiment the method comprises:

obtaining the input audio signal via the one or more microphones of the audio device.

[0050] The method according to the present invention may be performed in real-time on an obtained input audio signal, i.e., the one or more optimized audio processing parameters may be determined in real-time during an online conversation between two participants without substantial delay. To facilitate carrying out the method in real-time the one or more user parameters is preferably obtained prior to obtaining the input audio signal or is available for the audio device to obtain the one or more user parameters in real-time, e.g., by having the one or more user parameters stored on the audio device or available online for the audio device to retrieve the one or more user parameters.

[0051] In an embodiment the method comprises:

obtaining the input audio signal by receiving the input audio signal from a far-end audio device.

[0052] The far-end audio device is another audio device communicatively connected to the audio device, the audio device according to the disclosure may then act as the near-end audio device.

[0053] In an embodiment the method comprises:

determining perceptual information based on the one or more user parameters, and

determining the error function, wherein the error function is a function of the target audio signal, the input audio signal, the perceptual information and the one or more audio processing parameters.

[0054] The perceptual information may be based on one or more characteristics of the target user, such as one or more masking thresholds, one or more audible ranges, hearing loss, or audio preferences. The perceptual information may be in the form of a scalar, a vector, or a matrix. Preferably, the perceptual information is in the form of a weight and is used to weight the perceptually weight the error function, thus, forming a perceptually weighted error function. The perceptual information may be in the form of a weighting matrix. By including perceptual information in the error function, it may allow for a higher degree of personalization when determining the one or more optimized audio processing parameters. The perceptual information may be viewed as an additional perceptual constrain.

[0055] In an embodiment the perceptual information comprises one or more of the following one or more masking thresholds, one or more audible ranges, hearing loss, and audio preferences.

[0056] In an embodiment the method comprises:

determining the error function, wherein the error function is determined as $d_{w,x}(x, f(y, h)) = (x - f(y, h))^T W(x - f(y, h))$, wherein x denotes the target audio signal, y denotes the input audio signal, h denotes the one or more audio processing parameters, and W denotes the determined perceptual information,

determining the one or more optimized audio processing parameters, wherein the one more optimized audio processing parameters are determined by solving the following constrained optimization problem $h^* = \min_h \ d_{w,x}(X, f(x, h))$ constrained by the perceptual constraint, wherein h^* denotes the one or more optimized audio processing parameters.

[0057] In the above a squared error function is presented, however, it is readily understood for the person skilled in the art that other error functions may be equally applicable. Other L-norm functions may also be applicable as error functions. Further any well-behaved function with derivatives defined, or smooth for backpropagation needed by stochastic gradient descent routine need in neural network training may suffice as an error function. The squared error function has the advantage of being simple, thus, with sufficient assumptions on the target audio signal a closed-form solution to minimizing the error function is obtainable, alternatively, the error function may be minimized via numerical approaches.

[0058] In an embodiment the method comprises

determining a perceptual constraint, wherein the perceptual constraint is determined as $d_{w,x}\big(X, f(x, h)\big) = \sum_{i=1}^{I} \sum_k d_{w,x,i,k} \leq \beta_{i,k}$, wherein I denotes one or more processing blocks through which the input audio signal is processed, k denotes a frequency scale, and $\beta_{i,k}$ denotes the perceptual constrain, wherein $\beta_{i,k}$ comprises frequency-dependent thresholds for the one or more processing blocks determined based on the one or more user parameters.

[0059] Consequently, a perceptual constrain is determined which is compatible with different processing blocks of a digital signal processing chain. The processing blocks may be a noise reducer, an echo canceller, a speech restorer, an equalizer, etc.. In the present context frequency dependent may be understood as an input audio signal being split into several frequency bin, each bin having its own perceptual constrain associated with it.

[0060] In an embodiment the one or more optimized audio processing parameter comprises one or more audio filter weights.

[0061] In an embodiment the method comprises:

obtaining the input audio signal from a training dataset comprising a plurality of audio signals

determining an error function, wherein the error function is determined as $J_P = \sum_{i=1}^{I} \sum_k H_{i,k}(x, f(y, h), g_{i,k})$, wherein I denotes one or more types of audio degradation, k denotes a frequency index, $g_{i,k}$ denotes the perceptual constrain for the one or more types of audio degradation determined based on the one or more user parameters, $H_{i,k}(.)$ denotes the sub-term loss associated with the one or more types of audio degradation and the frequency index, x denotes the target audio signal, y denotes the input audio signal, and h denotes the one or more audio processing parameters,

determining the one or more optimized audio processing parameters, wherein the one more optimized audio processing parameters are determined by solving an optimization problem $h^* = \min_h \ J_p$, wherein h^* denotes the one or more optimized audio processing parameters.

[0062] In an embodiment the one or more optimized audio processing parameters is determined by stochastic gradient descent.

[0063] Although stochastic gradient descent is mentioned as one preferred solution, the present disclosure is not limited to this. Other numerical approaches may be equally applicable. For example, different types of gradient descent may be equally applicable as numerical solvers, alternatively derivative free methods may also be used, such as Bayesian optimization, pattern search, genetic algorithms, etc.. Furthermore, in some embodiments sufficient assumptions may be made to derive a closed-form solution. Such assumption may comprise assuming a Gaussian distribution on noise

or a specific distribution on speech, e.g., a super-gaussian.

[0064] In an embodiment the one or more optimized audio processing parameter comprises one or more weights for a machine learning model.

[0065] In an embodiment the method comprises:

applying the one or more optimized audio processing parameters to the input audio signal to generate an output audio signal, and

outputting the output audio signal.

[0066] The one or more optimized audio processing parameters may be applied to the input audio signal in a plethora of manners. The one or more optimized audio processing parameters may be one or more audio filter parameters of one or more audio filters, thus, by applying the audio filters to the input audio signal, the one or more optimized audio parameters are applied to the input audio signal. The one or more optimized audio processing parameters may be one or more neural network weights within a neural network, thus, by applying the neural network to the input audio signal, the one or more optimized audio parameters are applied to the input audio signal.

[0067] Outputting the output audio signal may comprise transmitting the output audio signal to an output transducer, such as a loudspeaker. Outputting the output audio signal may comprise transmitting the output audio signal to another audio device. Outputting the output audio signal may comprise transmitting the output audio signal to further processing circuitry or further digital signal processing blocks.

[0068] According to a second aspect of the invention there is provided an audio device comprising an interface, a speaker, and one or more microphones, a processor and a memory, wherein the audio device is configured to:

obtain an input audio signal,

obtain one or more user parameters indicative of one or more characteristics of a target user,

determine an error function, wherein the error function is a function of a target audio signal, the input audio signal and the one or more audio processing parameters,

determine a perceptual constraint based on the one or more user parameters,

determine one or more optimized audio processing parameters by minimizing the error function constrained by the perceptual constrain, and wherein the one or more audio processing parameters are the optimization variables, and

output the one or more optimized audio processing parameters.

[0069] The audio device of the second aspect may be configured to carry the steps of the method according to the first aspect of the present disclosure. The steps of the method according to the first aspect of the present disclosure may form a computer-implemented method, where the steps of the method is carried out by a computer.

BRIEF DESCRIPTION OF THE DRAWINGS

[0070] The above and other features and advantages of the present disclosure will become readily apparent to those skilled in the art by the following detailed description of example embodiments thereof with reference to the attached drawings, in which:

Fig. 1 is a block diagram of an audio device according to the present disclosure.

Fig. 2 is a conceptual graph of the amplitude of the attenuation of different distortion types as a function of filter parameters.

Fig. 3 is a flow chart for carrying out the method according to an embodiment of the present disclosure.

DETAILED DESCRIPTION

[0071] Various example embodiments and details are described hereinafter, with reference to the figures when relevant. It should be noted that the figures may or may not be drawn to scale and that elements of similar structures or functions

are represented by like reference numerals throughout the figures. It should also be noted that the figures are only intended to facilitate the description of the embodiments. They are not intended as an exhaustive description of the disclosure or as a limitation on the scope of the disclosure. In addition, an illustrated embodiment needs not have all the aspects or advantages shown. An aspect or an advantage described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced in any other embodiments even if not so illustrated, or if not so explicitly described.

[0072] Referring initially to figure 1 which depicts a block diagram of an audio device 10 according to the present disclosure. The audio device 10 may be a speakerphone, a headset, one or more earbuds, a computer, a tablet, or a smart phone. The audio device 10 may be used for a conference and/or a meeting between two or more parties being remote from each other. Optionally, the audio device 10 may be communicatively connected to a server device 20. Optionally, the audio device 10 may be communicatively connected to a far-end communication device 30. The communication device 30 may be seen as a communication device used by one or more far-end users to communicate with the one or more users of the audio device 10 via a network such as global network, e.g., the internet, and/or a local network. Optionally, the audio device 10 may be communicatively connected to an electronic device 40. The electronic device 40 may be a smartphone, a smart-watch, a conference hub, a smart-tv, smart-speakers, a tablet, a computer, such as a laptop computer or PC, or a tablet computer. In other words, the electronic device 40 may be a user device configured to communicate with the audio device 10. The electronic device 40 may be provided with a user interface allowing for a user of the electronic device 40 to control one or more settings of the audio device. The electronic device 40 may facilitate the transfer of user inputs from the electronic device to the audio device 10.

[0073] The audio device 10 comprises an interface 11. The interface 11 allows for the audio device 10 to interface with other devices. The interface 11 may comprise a wireless transceiver allowing for the interface 11 to set-up bidirectional communication lines with other devices. The interface 11 may provide for a wired connected with other devices. The audio device 10 comprises a speaker 12. The speaker 12 is configured to output an audio signal. The speaker 12 may output an audio signal received from another device, such as the server device 20, the far-end communication device 30, and/or the electronic device 40. The speaker 12 may output an audio signal processed by a processor 14 of the audio device 10. The audio device 10 comprises one or more microphones 13. The one or more microphones 13 may form a microphone array. The one or more microphones 13 may be configured to obtain an input audio signal. The one or more microphones 13 may be configured to obtain an input audio signal associated with a user of the audio device 10. The audio device 10 comprises a processor 14. The processor 14 may be any unit capable of executing logic routines such as lines of code, software programs, etc. The audio device comprises a memory 15. The memory 15 may be one or more of a buffer, a flash memory, a hard drive, a removable media, a volatile memory, a non-volatile memory, a random access memory (RAM), or other suitable device. In a typical arrangement, the memory 15 may include a non-volatile memory for long term data storage and a volatile memory that functions as system memory for the processor 14. The memory 15 may exchange data with the processor 14 over a data bus. Control lines and an address bus between the memory 15 and the processor 14 may be present. The memory 15 may be considered a non-transitory computer readable medium.

[0074] The audio device 10 is configured to obtain an input audio signal. The audio device 10 may obtain the input audio signal by using the processor 14. The input audio signal may be obtained by the one or more microphones 13 of the audio device 10. The input audio signal may be obtained from a training dataset comprising a plurality of audio signals. The input audio signal may be received from the far-end 30. The input audio signal may be obtained by retrieving the input audio signal from the memory 15 of the audio device 10.

[0075] The audio device 10 is configured to obtain one or more user parameters. The audio device 10 may obtain the one or more user parameters by using the processor 14. The one or more user parameters are indicative of one or more characteristics of a target user. The one or more user parameters may be obtained by receiving one or more inputs from a target user. The one or more inputs may be received via the electronic device 40, the far-end device 30, or the server device 20. The one or more user parameters may be obtained by retrieving the one or more user parameters from the memory 15 of the audio device.

[0076] The audio device 10 is configured to determine an error function. The audio device 10 may determine the error function by using the processor 14. The error function is a function of a target audio signal, the input audio signal and one or more audio processing parameters.

[0077] The audio device 10 is configured to determine a perceptual constrain. The audio device 10 may determine the perceptual constrain by using the processor 14. The perceptual constraint may be any weight and/or threshold based on the one or more user parameters.

[0078] The audio device 10 is configured to determine one or more optimized audio processing parameters. The audio device 10 may determine the one or more optimized audio processing parameters by using the processor 14. The one or more optimized audio processing parameters are determined by minimizing the error function constrained by the perceptual constrain and using the one or more audio processing parameters as the optimization variables. The audio device 10 may be configured to determine the one or more optimized audio processing parameters by numerically

solving the optimization problem. The audio device 10 may be configured to determine the one or more optimized audio processing parameters by deriving a closed form solution to the optimization problem. The audio device 10 may be configured to determine one or more assumption to derive a closed form solution to the optimization problem.

[0079] The audio device 10 is configured to output the one or more optimized audio processing parameters. The audio device 10 may output the one or more optimized audio processing parameters by using the processor 14. The one or more optimized audio processing parameters may be outputted to an audio filter of the audio device 10. The one or more optimized audio processing parameters may be outputted to a neural network of the audio device 10. The audio device 10 may output the one or more optimized audio processing to the electronic device 40, the far-end device 30, and/or the server device 20.

[0080] Referring to Fig. 2 which shows a conceptual graph 50 of the amplitude of the attenuation of different distortion types in arbitrary units as a function of filter parameters in arbitrary units. The graph is merely conceptual in nature and is meant to elucidate the disclosure in further detail. The graph depicts three different curves, namely the noise attenuation curve 52, speech restoration curve 53 and a masking threshold 51. In prior art audio devices when processing received audio signals the end output, is normally the result of a long series of trade-off to try to achieve an optimal result. The graph 50 depicts one example of a dilemma leading to a trade-off. The trade-off occurs because the optimal audio processing parameters for achieving maximum noise attenuation are not necessarily the same audio processing parameters achieving maximum speech restoration. Consequently, a trade-off needs to be made, as it is not possible to achieve the optimal result for both noise attenuation and speech restoration. One simple way to then select the audio processing parameters, is to determine an overall distortion attenuation and choosing the audio parameters leading to the lowest overall distortion attenuation. However, such calculation of overall distortion attenuation assumes that each distortion type should be equally weighted, even though this is not necessarily the case. For example, some people are more sensitive to noise than others, hence, determining the audio processing parameters based on a general approach without considering personal preferences may lead to some of the trade-offs when selecting audio processing parameters giving a worse result for some if not all users. A further aspect which is not taken into the account is that everyone has a different hearing profile, thus, leading to different masking thresholds. In prior art audio devices, the same hearing profile is assumed for everyone. Assuming each hearing profile is the same may also lead to severe errors, as even though an objective improvement in the distortion attenuation is achieved, such an improvement may in some cases not be perceivable if the changes taking place are below a masking threshold of the user. Consequently, prior art audio devices suffer under having a too generalist approach in determining audio processing parameters, such an approach may fail to take into consideration the listening preferences and/or the hearing profile of the user.

[0081] Referring to Fig. 3 which shows a flow chart 100 for carrying out the method according to an embodiment of the present disclosure. The flow chart 100 will be described in the following in relation to two different methods of carrying out the present disclosure, namely, a method focused on digital signal processing, and a method focused on a method for training a machine learning model, such as a neural network.

Digital signal processing

[0082] The following presented method focuses on digital signal processing, e.g., determining one or more filter parameters and/or weights for an audio filter of an audio device. However, the presented method may be equally applicable as a method for training a machine learning model.

[0083] The method comprises a first step 101 comprising obtaining an input audio signal. The input audio signal may be obtained from a microphone associated with an audio device carrying out the method. The input audio signal may be obtained by receiving it from a far-end device, such as another audio device, an electronic device, or a server device. The input audio signal may be obtained by retrieving it from a memory of the audio device carrying out the method.

[0084] The method comprises a second step 102 comprising obtaining one or more user parameters. The one or more user parameters are indicative of one or more characteristics of a target user. The one or more user parameters may be obtained by receiving it from a far-end device, such as another audio device, an electronic device, or a server device. The one or more user parameters may be obtained by retrieving it from a memory of the audio device carrying out the method. The one or more user parameters may have been obtained by performing one or more test on the target user, e.g., a hearing test, a listening test, etc. The one or more user parameters may have been obtained by receiving one or more inputs from the target user, the one or more inputs may have been given directly to the audio device carrying out the method or via an external device communicatively connected to the audio device.

[0085] The method optionally comprises a third step 103 comprising determining perceptual information. The perceptual information may be determined as a weighting matrix W. As an example, W can be an N-dimensional diagonal perceptual weighting matrix, such as

$$W = \begin{bmatrix} T(0) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & T((N-1)\omega_0) \end{bmatrix}$$

[0086] Such perceptual weighting matrix are already known from low-rate analysis-by-synthesis speech coders, e.g., CELP. However, the known perceptual weighting matrices are not personalized and assume a common hearing profile. Here $\omega_0 = 2\pi/N$ and $T(\omega_0)$ denotes a frequency response, e.g., the one used in CELP coded is used for shaping the noise distortion in the frequency domain. In the present disclosure, the frequency response may be personalized based on the one or more user parameters and utilized to shape distortion based on the hearing profile and/or listening preferences of the target user. *W* may be seen as a masking threshold tolerance or as a set of frequency-dependent shaping filters in time *T(.)* as written in the above *W* matrix. The role of the perceptual information is to capture the user preference with regards to various distortion types via thresholding or shaping various distortion types (echo annoyance, residual noise, reverberation) towards a user's preference. *W* may be a function of frequency, and thus may relate to a filter response, hence it is possible to shape distortions to make them inaudible/masked which may be preferred by the end user, thus, achieving a personalized result.

[0087] The method comprises a fourth step 104 comprising determining an error function. The error function may be determined as a simple quadratic error function

$$d_{,x}(x, \hat{x}) = (x - \hat{x})^T (x - \hat{x})$$

where x denotes a target audio signal, and $\hat{x}$ denotes an estimator. If the perceptual information has been determined, the error function may be perceptually weighted and determined as

$$d_{,x}(x, \hat{x}) = (x - \hat{x})^T W (x - \hat{x})$$

[0088] Where W denotes the perceptual information, e.g., in the form of a weighting matrix. Furthermore, the estimator x may be rewritten to

$$\hat{x} = f(y, h)$$

where *h* denotes one or more audio processing parameters, and y denotes an input audio signal, and the function *f*(.) represents the modulation the audio processing parameters perform on the input audio signal. The one or more audio processing parameters may be non-personalized audio processing parameters, e.g., generic audio processing parameters set by the manufacturer of the audio device. The one or more audio processing parameters may be one or more previously determined audio processing parameters, such as one or more previously determined optimized audio processing parameters. The rewritten estimator may be inserted in the error function, and thus the error function can be expressed as

$$d_{w,x}(x, f(y, h)) = (x - f(y, h))^T W (x - f(y, h))$$

[0089] The method comprises a fifth step 105 comprising determining a perceptual constraint. The perceptual constraint is determined based on the one or more user parameters of the target user. The perceptual constraint may be determined from one or more listening tests conducted on the target user. The listening tests may be used for determining one or more listening preferences of the target user regarding different types of distortions, such as noise, speech distortion, dereverberation artifact, echo annoyance tolerance. As an example, the subjective listening test may comprise the target user rating different sound clips with MOS scores. From the MOS scores of the listening test conducted on the target user, the preferences of the target user may be determined. The preferences of the user may then be converted to one or more perceptual constraints. Thus, the perceptual constraint can be expressed as

$$g(a_{i,k}) = \beta_{i,k}$$

where $a_{i,k}$ denotes the one or more user parameters, $\beta_{i,k}$ denotes the one or more perceptual constraints, and $g(.)$ may be any function for converting the $a_{i,k}$ to $\beta_{i,k}$. In one embodiment the one or more user parameters may directly be determined as the perceptual constraint. The subscript i denotes different audio distortion types, and the subscript k denotes different frequency ranges and/or bins. Although it is stated k denotes different frequency ranges and/or bins, it should not be construed in a limiting manner, and k may alternatively denote different ranges and/or bins in the bark scale or mel scale. Hence, $\beta_{i,k}$ may be viewed as one or more frequency-dependent thresholds associated with one or more different audio distortion types.

[0090] The choice of g(.) may be heuristic or based on some inspection concluded from the listening test outcome. In one example, the user is asked to add inputs about the max noise attenuation depth G_{min}, e.g., in decibels, that the user prefers depending on various noisy audio examples. G_{min} is then the one or more user parameters $a_{i,k}$. G_{min} may then be converted to an over/under-subtraction factor $\beta_{i,k}$ using an exponential function to get into linear domain to emphasize/de-emphasize on how much noise needs to be attenuated in the constrained optimization formulation. In particular, this $\beta_{i,k}$ can be used to limit the gain G_{pre} calculated via digital signal processing or a machine learning model to get the ultimate output gain as $G_{post} = \min(G_{pre}, G_{min})$.

[0091] The method comprises a sixth step 106 comprising determining one or more optimized audio processing parameters. The one or more optimized processing parameters h^* are determined by minimizing the error function constrained by the perceptual constrain, and wherein the one or more audio processing parameters are the optimization variables. This may be formulated as

$$h^* = \arg\min_{\boldsymbol{h}} \; d_{w,x}(x, f(y,h)) \qquad subj: d_{w,x}\big(X, f(x,h)\big) = \sum_{i=1}^{I} \sum_k d_{w,x,i,k} \leq \beta_{i,k}$$

where I denotes one or more processing blocks through which the input audio signal is processed, e.g., an echo canceller, a noise suppressor, etc.. The above present optimization problem may either be solved numerically, or a closed-form solution may be derived by making appropriate assumptions on the undesired distortions. By solving the presented optimization problem, one will arrive at a set of optimized filter weights for each of the processing blocks through which the input audio signal is processed, furthermore, the set of optimized filter weights takes into account the listening preferences and/or the hearing profile of the target user.

[0092] The method comprises a seventh step 107 comprising outputting one or more optimized audio processing parameters. The one or more optimized audio processing parameters may be outputted to an audio filter of the audio device performing the method. The one or more optimized audio processing parameters may be outputted to another device, such as another audio device, an electronic device, or a server device.

[0093] The method optionally comprises an eight step 108 comprising processing the input audio signal using the one or more determined optimized processing parameters.

[0094] The method optionally comprises a ninth step 109 comprising outputting the processed input audio signal.

[0095] One or more steps of the presented method may be repeated one or more times, thus, forming an iterative method, e.g., the fourth step 104, the sixth step 106 and the seventh step 107 may be repeated for each new determined one or more optimized audio processing parameters, until it is observed that the new determined one or more optimized audio processing parameters starts to converge or have converged. Other stopping criterions for the method may comprise repeating the method for a pre-determined number of iterations or continue until the changes in the determined optimized audio processing parameters are below a pre-defined tolerance value. The second step 102, the third step 103, and the fifth step 105 need not be repeated. The steps of the methods need not be carried in the presented order, e.g., the second step 102, the third step 103, and the fifth step 105 may be carried out prior to the other steps such as the fourth step 104, the sixth step 106 and the seventh step 107. In one embodiment, the steps second step 102, the third step 103, and the fifth step 105 are carried out prior to an online-conference or when initiating the online-conference, and the fourth step 104, the sixth step 106 and the seventh step 107 are carried out in-real time during the online conference. All the steps or one or more of the steps of the presented method may be carried out in-real time. The second step 102, the third step 103, and the fifth step 105 may be repeated at different times to reflect changes in the target user's listening preferences or changes in the hearing profile of the target user.

Training of a machine learning model

[0096] The following presented method focuses on training of a machine learning model, e.g., determining one or more weights of a neural network. However, the presented method may be equally applicable as a method for determining one or more filter parameters.

[0097] The method comprises a first step 101 comprising obtaining an input audio signal. The input audio signal may be obtained from a microphone associated with an audio device carrying out the method. The input audio signal may

be obtained by receiving it from a far-end device, such as another audio device, an electronic device, or a server device. The input audio signal may be obtained by retrieving it from a memory of the audio device carrying out the method. The obtained input audio signal may be part of a larger set of audio signals.

[0098] The method comprises a second step 102 comprising obtaining one or more user parameters. The one or more user parameters are indicative of one or more characteristics of a target user. The one or more user parameters may be obtained by receiving it from a far-end device, such as another audio device, an electronic device, or a server device. The one or more user parameters may be obtained by retrieving it from a memory of the audio device carrying out the method. The one or more user parameters may have been obtained by performing one or more test on the target user, e.g., a hearing test, a listening test, etc. The one or more user parameters may have been obtained by receiving one or more inputs from the target user, the one or more inputs may have been given directly to the audio device carrying out the method or via an external device communicatively connected to the audio device.

[0099] The method comprises a fourth step 104 comprising determining an error function. The error function J_P may be determined as perceptually motivated loss function

$$J_P = \sum_{i=1}^{I} \sum_{k} H_{i,k}(x, f(y, \boldsymbol{h}), i, g_{i,k})$$

[0100] I denotes one or more types of audio degradation, k denotes a frequency index, $g_{i,k}$ denotes the perceptual constraint for the one or more types of audio degradation determined based on the one or more user parameters, $H_{i,k}()$ denotes the sub-term loss associated with the one or more types of audio degradation and the frequency index, x denotes the target audio signal, y denotes the input audio signal, and *h* denotes the one or more audio processing parameters. $H_{i,k}()$ may denote any function defined as the difference between the clean signal *x* and the enhanced signals *f*(*y,* ***h***), e.g., mean square error (MSE), other psychoacoustical-motivated error criteria, or norm or non-linear mapping from x to *f*(*y,* ***h***). The function *f*(.) represents the filtering/convolution the audio processing parameters perform on the input audio signal.

[0101] The presented perceptual constraint $g_{i,k}$ differs from that of $\beta_{i,k}$, as $\beta_{i,k}$ was defined as one or more threshold values, where $g_{i,k}$ in the presented method is given as one or more frequency dependent and distortion dependent weights or shaping filter frequency responses (denoted by T(.) in the earlier example). In other words, $g_{i,k}$ is a weight determining the relative importance of each of the one or more sub-term losses. However, both are in the present disclosure viewed as perceptual constraints as both $g_{i,k}$ and $\beta_{i,k}$ lead to solution space for the error function which is constrained perceptually according to the listening preferences and/or the hearing profile of the target user.

[0102] As one specific example, the parameterized loss function presented by S. Braun, M. Luis Valero, "Task Splitting for DNN-based acoustic echo and noise removal", published at IWAENC 2022, Bamberg, Germany may be used. There the loss function is given by

$$L = L(\hat{S}, S) + lL_{asym}(\hat{s}, s)$$

which is a combination of two individual losses: one symmetric loss combining magnitude and phase-sensitive costs:

$$L(\hat{S}, S) = \sum_{k,n} \left| |\hat{S}|^c e^{j\varphi_{\hat{S}}} - |S|^c e^{j\varphi_S} \right|^2 + g_0 \left| |\hat{S}|^c - |S|^c \right|^2$$

and an asymmetric loss

$$L_{asym}(\hat{S}, S) = \sum_{k,n} max\{|\hat{S}|^c - |S|^c, 0\}^2$$

where g_0 is a preference weighting coefficient reflecting how much the user is an echo leak hater hence penalizing more via emphasizing the asymmetric loss sub-term, c is a compression applied to the spectral magnitude, *k* and *n* are the frequency and time indices, respectively, S and $\hat{S}$ are the desired target and estimated speech signals, respectively. The idea would be to associate the perceptual constrain $g_{i,k}$ to the parametrizations in the above examples, e.g., c, g_0 and *l*. Thereby, providing a perceptually motivated loss function.

[0103] The method comprises a fifth step 105 comprising determining a perceptual constraint. The perceptual constraint

is determined based on the one or more user parameters of the target user. The perceptual constrain may be determined from one or more listening tests conducted on the target user. The listening tests may be used for determining one or more listening preferences of the target user regarding different type of distortions, such as noise attenuation, speech distortion, dereverberation artifact, echo annoyance tolerance. As an example, the subjective listening test may comprise the target user rating different sound clips with MOS scores. From the MOS scores of the listening test conducted on the target user, the preferences of the target user may be determined. The preferences of the user may then be converted to one or more perceptual constraints. Thus, the perceptual constraint can be expressed as

$$j(a_{i,k}) = g_{i,k}$$

where $a_{i,k}$ denotes the one or more user parameters, $g_{i,k}$ denotes the one or more perceptual constrains, and $j(.)$ may be any function for converting $a_{i,k}$ to $g_{i,k}$. In one embodiment the one or more user parameters may directly be determined as the perceptual constraint. The subscript i denotes different audio distortion types, and the subscript k denotes different frequency ranges and/or bins. Although it is stated k denotes different frequency ranges and/or bins, it should not be construed in a limiting manner, and k may alternatively denote different ranges and/or bins in the bark scale or mel scale. In one example, j(.) is a compression/expansion function log(.)/exp(.) or any generalized compression function with exponent c, c not being limited to be an integer. The method comprises a sixth step 106 comprising determining one or more optimized audio processing parameters. The one or more optimized processing parameters h^* are determined by minimizing the error function constrained by the perceptual constraint, and wherein the one or more audio processing parameters are the optimization variables. This may be formulated as

$$h^* = \min_h \ J_p$$

where h^* denotes the one or more optimized audio processing parameters. In the present example, where the method is focused on training a machine learning model, the machine learning model may be targeted to find the optimal weights found via minimizing J_p. The above present optimization problem may be solved numerically, e.g., by stochastic gradient descent. By solving the presented optimization problem, one will arrive at a set of optimal weights for the machine learning model.

[0104] The method comprises a seventh step 107 comprising outputting one or more optimized audio processing parameters. The one or more optimized audio processing parameters may be outputted to another device, such as another audio device, an electronic device, or a server device.

[0105] The method optionally comprises an eight step 108 comprising processing the input audio signal using the one or more determined optimized processing parameters.

[0106] The method optionally comprises a ninth step 109 comprising outputting the processed input audio signal.

[0107] One or more steps of the presented method may be repeated one or more times, thus, forming an iterative method, e.g., the first step 101, the fourth step 104, the sixth step 106 and the seventh step 107 may be repeated different input audio signals, until it is observed that the new determined one or more optimized audio processing parameters starts to converge or have converged. The second step 102, the third step 103, and the fifth step 105 need not be repeated. The steps of the methods need not be carried in the presented order, e.g., the second step 102, the third step 103, and the fifth step 105 may be carried out prior to the other steps such as the fourth step 104, the sixth step 106 and the seventh step 107. In one embodiment, the steps second step 102, the third step 103, and the fifth step 105 are carried out prior to an online-conference or when initiating the online-conference, and the fourth step 104, the sixth step 106 and the seventh step 107 are carried out in-real time during the online conference. All the steps or one or more of the steps of the presented method may be carried out in-real time. The second step 102, the third step 103, and the fifth step 105 may be repeated to reflect changes in the target user's listening preferences or changes in the hearing profile of the target user.

[0108] The use of the terms "first", "second", "third" and "fourth", "primary", "secondary", "tertiary" etc. does not imply any particular order but are included to identify individual elements. Moreover, the use of the terms "first", "second", "third" and "fourth", "primary", "secondary", "tertiary" etc. does not denote any order or importance, but rather the terms "first", "second", "third" and "fourth", "primary", "secondary", "tertiary" etc. are used to distinguish one element from another. Note that the words "first", "second", "third" and "fourth", "primary", "secondary", "tertiary" etc. are used here and elsewhere for labelling purposes only and are not intended to denote any specific spatial or temporal ordering.

[0109] Furthermore, the labelling of a first element does not imply the presence of a second element and vice versa.

[0110] It is to be noted that the word "comprising" does not necessarily exclude the presence of other elements or steps than those listed.

[0111] It is to be noted that the words "a" or "an" preceding an element do not exclude the presence of a plurality of

such elements.

[0112] It should further be noted that any reference signs do not limit the scope of the claims, that the example embodiments may be implemented at least in part by means of both hardware and software, and that several "means", "units" or "devices" may be represented by the same item of hardware.

[0113] The various example methods, devices, and systems described herein are described in the general context of method steps processes, which may be implemented in one aspect by a computer program product, embodied in a computer-readable medium, including computer-executable instructions, such as program code, executed by computers in networked environments. A computer-readable medium may include removable and non-removable storage devices including, but not limited to, Read Only Memory (ROM), Random Access Memory (RAM), compact discs (CDs), digital versatile discs (DVD), etc. Generally, program modules may include routines, programs, objects, components, data structures, etc. that perform specified tasks or implement specific abstract data types. Computer-executable instructions, associated data structures, and program modules represent examples of program code for executing steps of the methods disclosed herein. The particular sequence of such executable instructions or associated data structures represents examples of corresponding acts for implementing the functions described in such steps or processes.

[0114] Although features have been shown and described, it will be understood that they are not intended to limit the claimed disclosure, and it will be made obvious to those skilled in the art that various changes and modifications may be made without departing from the spirit and scope of the claimed disclosure. The specification and drawings are, accordingly, to be regarded in an illustrative rather than restrictive sense. The claimed disclosure is intended to cover all alternatives, modifications, and equivalents.

Claims

1. A method for determining one or more personalized audio processing parameters for an audio device comprising an interface, a speaker, and one or more microphones, the method comprising:

   obtaining an input audio signal,
   obtaining one or more user parameters indicative of one or more characteristics of a target user,
   determining an error function, wherein the error function is a function of a target audio signal, the input audio signal and one or more audio processing parameters,
   determining a perceptual constraint based on the one or more user parameters,
   determining one or more optimized audio processing parameters by minimizing the error function constrained by the perceptual constrain, and wherein the one or more audio processing parameters are the optimization variables, and
   outputting the one or more optimized audio processing parameters.

2. A method according to claim 1, wherein the one or more user parameters comprises physiological information regarding the target user, such as gender and/or age.

3. A method according to any of the preceding claims, wherein the one or more user parameters comprises a result of a hearing test carried out on the target user.

4. A method according to any of the preceding claims, comprising:
   obtaining the input audio signal via the one or more microphones of the audio device.

5. A method according to any of the preceding claims, comprising:

   determining perceptual information based on the one or more user parameters, and
   determining an error function, wherein the error function is a function of the target audio signal, the input audio signal, the perceptual information and the one or more audio processing parameters.

6. A method according to claim 5, wherein the perceptual information comprises one or more of the following one or more masking thresholds, one or more audible ranges, hearing loss, and audio preferences.

7. A method according to any of claims 5 or 6, comprising:

   determining an error function, wherein the error function is determined as $d_{w,x}(x, f(y, h)) = (x - f(y, h))^T W(x - f(y, h))$, wherein x denotes the target audio signal, y denotes the input audio signal, h denotes the one or more

audio processing parameters, and W denotes the determined perceptual information,
determining the one or more optimized audio processing parameters, wherein the one more optimized audio processing parameters are determined by solving the following constrained optimization problem $h^* = \min_h \ d_{w,x}(X, f(x,h))$ constrained by the perceptual constraint, wherein h^* denotes the one or more optimized audio processing parameters.

8. A method according to claim 7, comprising:
determining a perceptual constraint, wherein the perceptual constraint is determined as $d_{w,x}\big(X, f(x,h)\big) = \sum_{i=1}^{I} \sum_k d_{w,x,i,k} \leq \beta_{i,k}$, wherein *I* denotes one or more processing blocks through which the input audio signal is processed, k denotes a frequency scale, and $\beta_{i,k}$ denotes the perceptual constrain, wherein $\beta_{i,k}$ comprises frequency-dependent thresholds for the one or more processing blocks determined based on the one or more user parameters.

9. A method according to any of the preceding claims, wherein the one or more optimized audio processing parameter comprises one or more audio filter weights.

10. A method according to any of claims 1 to 3, and 5 to 6, comprising:

obtaining the input audio signal from a training dataset comprising a plurality of audio signals.
determining an error function, wherein the error function is determined as $J_P = \sum_{i=1}^{I} \sum_k H_{i,k}(x, f(y,h), g_{i,k})$, wherein I denotes one or more types of audio degradation, *k* denotes a frequency index, $g_{i,k}$ denotes the perceptual constrain for the one or more types of audio degradation determined based on the one or more user parameters, $H_{i,k}()$ denotes the sub-term loss associated with the one or more types of audio degradation and the frequency index, *x* denotes the target audio signal, y denotes the input audio signal, and *h* denotes the one or more audio processing parameters,
determining the one or more optimized audio processing parameters, wherein the one more optimized audio processing parameters are determined by solving an optimization problem $h^* = \min_h \ J_p$, wherein h^* denotes the one or more optimized audio processing parameters.

11. A method according to claim 10, wherein the one or more optimized audio processing parameters is determined by stochastic gradient descent.

12. A method according to any of claim 10-11, wherein the one or more optimized audio processing parameter comprises one or more weights for a machine learning model.

13. A method according to any of the proceeding claims, comprising:

applying the one or more optimized audio processing parameters to the input audio signal to generate an output audio signal, and
outputting the output audio signal.

14. An audio device comprising an interface, a speaker, and one or more microphones, a processor and a memory, wherein the audio device is configured to:

obtain an input audio signal,
obtain one or more user parameters indicative of one or more characteristics of a target user,
determine an error function, wherein the error function is a function of a target audio signal, the input audio signal and the one or more audio processing parameters,
determine a perceptual constraint based on the one or more user parameters,
determine one or more optimized audio processing parameters by minimizing the error function constrained by the perceptual constrain, and wherein the one or more audio processing parameters are the optimization variables, and

output the one or more optimized audio processing parameters.

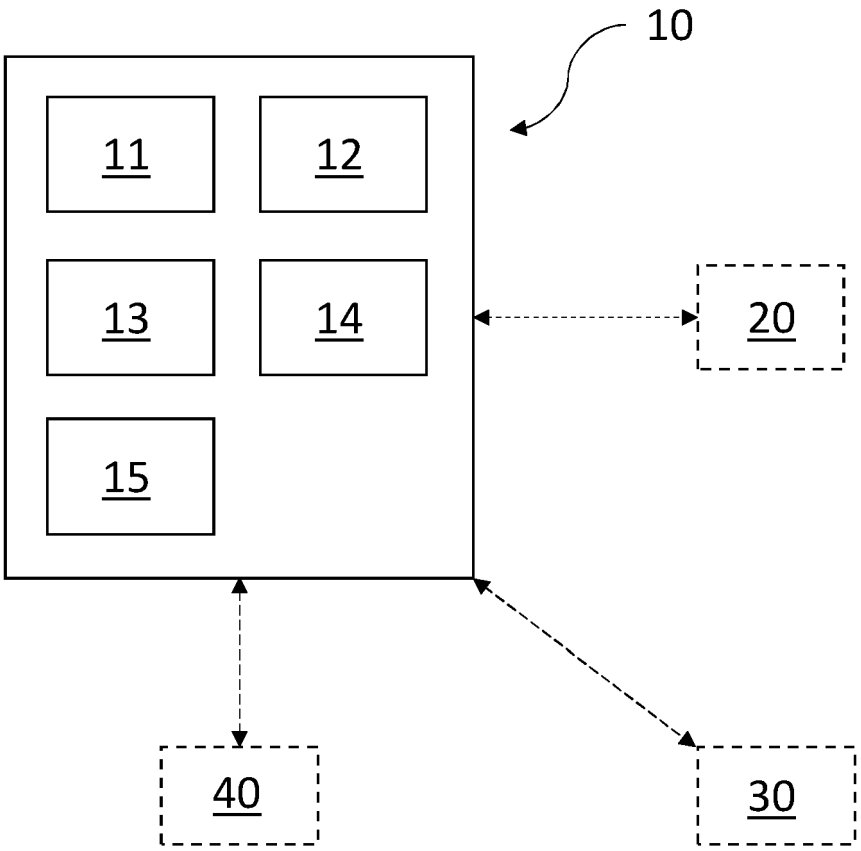

Fig. 1

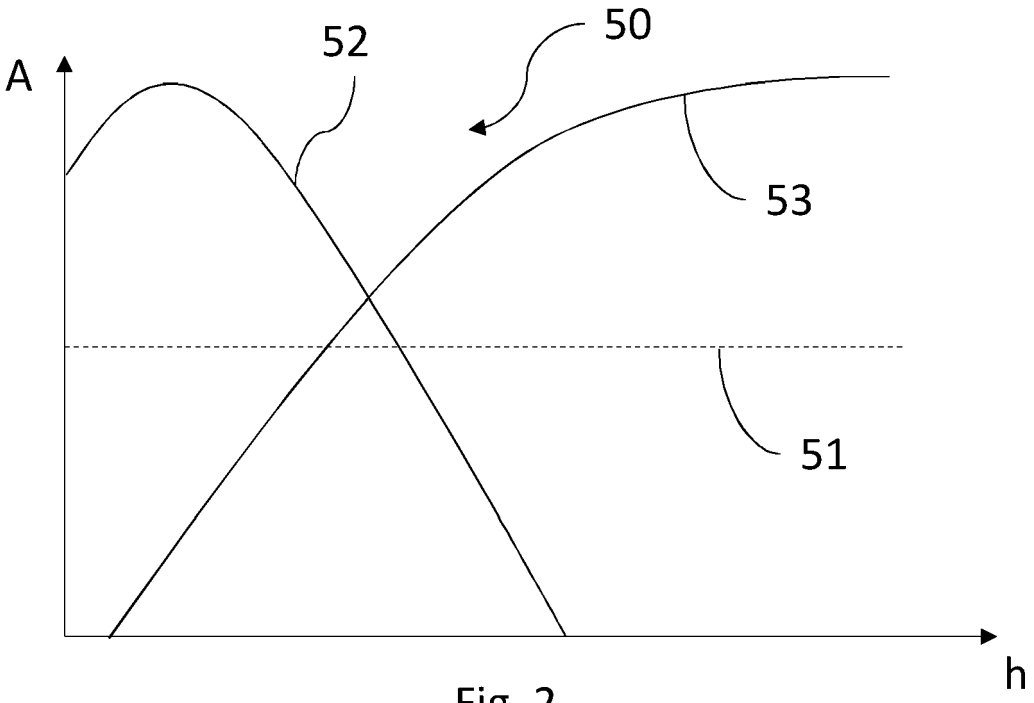

Fig. 2

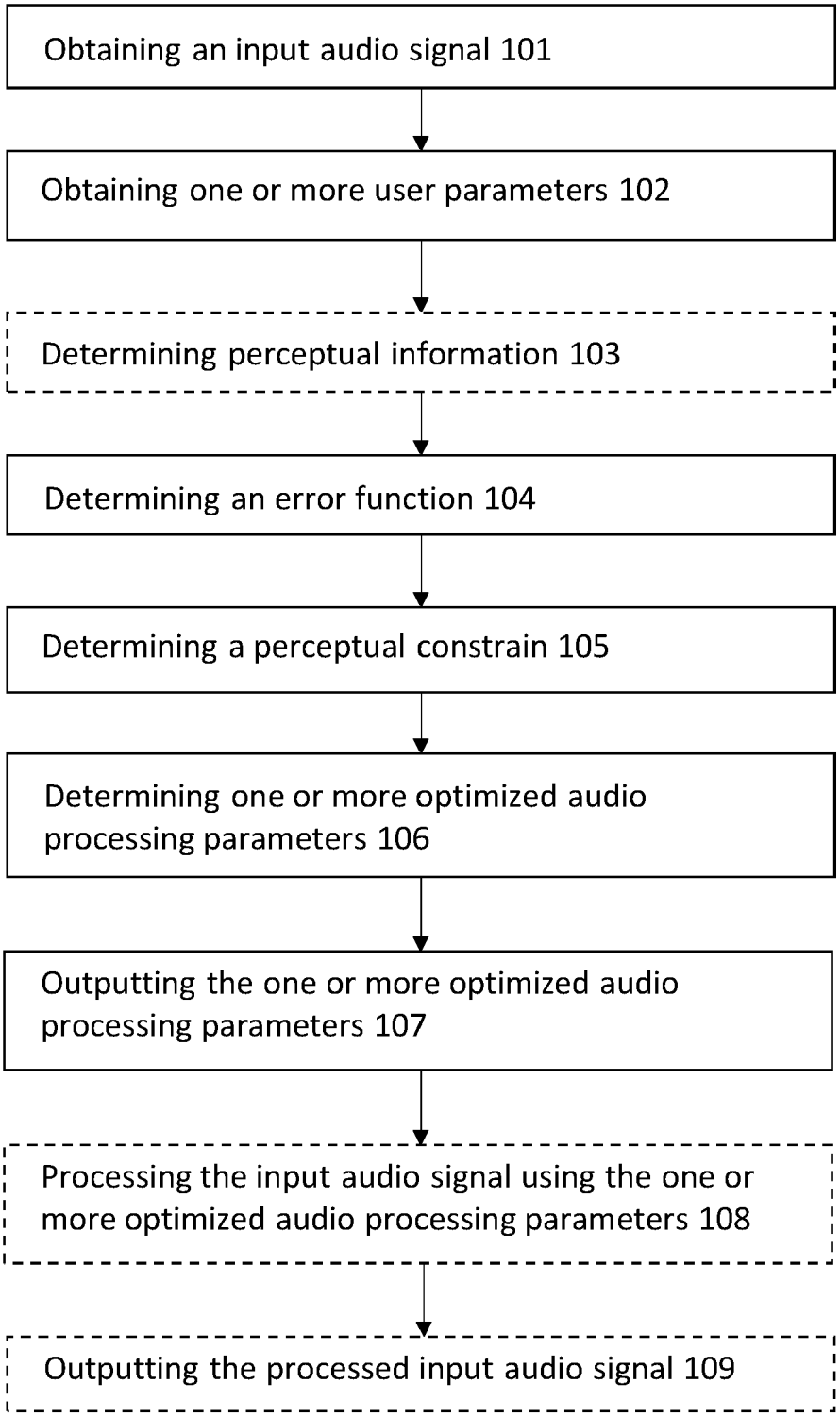
Obtaining an input audio signal 101
Obtaining one or more user parameters 102
Determining perceptual information 103
Determining an error function 104
Determining a perceptual constrain 105
Determining one or more optimized audio processing parameters 106
Outputting the one or more optimized audio processing parameters 107
Processing the input audio signal using the one or more optimized audio processing parameters 108
Outputting the processed input audio signal 109

Fig. 3

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 22 19 6043

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,D	WO 2020/016440 A1 (MIMI HEARING TECH GMBH [DE]) 23 January 2020 (2020-01-23) * figures 9, 22 * * paragraphs [0010], [0015], [0080] - [0082], [0092], [0103], [0107] *	1-14	INV. G10L25/27 ADD. G10L21/02 G10L25/30
A	US 2005/129262 A1 (DILLON HARVEY [AU] ET AL) 16 June 2005 (2005-06-16) * figures 2, 3 * * paragraphs [0118] - [0177], [0183], [0184] *	1-14	H04R25/00 H04R5/04
A	ZEHAI TU ET AL: "Optimising Hearing Aid Fittings for Speech in Noise with a Differentiable Hearing Loss Model", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 8 June 2021 (2021-06-08), XP081986836, * figures 1, 2 * * section 2 *	1-14	

TECHNICAL FIELDS SEARCHED (IPC)

G10L
H04R
H04S

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	27 January 2023	Tilp, Jan

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 22 19 6043

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.

27-01-2023

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2020016440 A1	23-01-2020	EP 3598440 A1	22-01-2020
		EP 3598441 A1	22-01-2020
		US 2020027467 A1	23-01-2020
		WO 2020016440 A1	23-01-2020
US 2005129262 A1	16-06-2005	US 2005129262 A1	16-06-2005
		US 2011202111 A1	18-08-2011

EPO FORM P0459

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

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 2020016440 A **[0005]**

Non-patent literature cited in the description

- **S. BRAUN ; M. LUIS VALERO.** Task Splitting for DNN-based acoustic echo and noise removal. *IWAENC,* 2022 **[0102]**