(51) **International Patent Classification:**
H04R 3/04 (2006.01)

(21) **International Application Number:**
PCT/EP2019/053898

(22) **International Filing Date:**
15 February 2019 (15.02.2019)

(25) **Filing Language:** English

(26) **Publication Language:** English

(71) **Applicant: HUAWEI TECHNOLOGIES CO., LTD.**
[CN/CN]; Huawei Administration Building, Bantian Longgang District, Shenzhen, Guangdong 518129 (CN).

(72) **Inventor; and**
(71) **Applicant** *(for US only)***: PANG, Liyun** [CN/DE]; c/o Huawei Technologies Duesseldorf GmbH, Riesstr. 25, 80992 Munich (DE).

(72) **Inventors: ADRIAENSEN, Fons**; Huawei Technologies Duesseldorf GmbH, Riesstr. 25, 80992 Munich (DE). **SCHLIEPER, Roman**; Gottfried Wilhelm Leibniz Universität, Hannover Welfengarten 1, 30167 Hannover (DE). **LI, Song**; Gottfried Wilhelm Leibniz Universität, Hannover Welfengarten 1, 30167 Hannover (DE).

(74) **Agent: KREUZ, Georg**; Huawei Technologies Duesseldorf GmbH, Riesstr. 25, 80992 Munich (DE).

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

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

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

(54) **Title:** METHOD AND APPARATUS FOR PROCESSING AN AUDIO SIGNAL BASED ON EQUALIZATION FILTER

FIG. 13
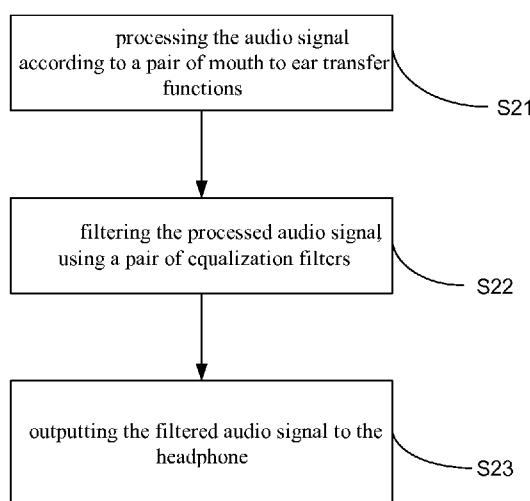
(57) **Abstract:** The invention relates to a method for processing an audio signal, the method comprising: processing the audio signal according to a pair of mouth to ear transfer functions, to obtain a processed audio signal; filtering the processed audio signal, using a pair of equalization filters, to obtain a filtered audio signal, wherein a parameter of the equalization filter is depends on an acoustic impedance of a headphone; and outputting the filtered audio signal to the headphone. According to this method to counteract the occlusion effect and to provide a natural perceived sound pressure.

METHOD AND APPARATUS FOR PROCESSING AN AUDIO SIGNAL BASED ON EQUALIZATION FILTER

Technical Field

5

The present invention relates to the field of audio signal processing and reproduction. More specially, the invention relates to a method for processing an audio signal based using an equalization filter and an apparatus for processing an audio signal based on equalization filter. The present invention also relates to a computer-readable storage medium.

10

Background

Headphones are a pair of small loudspeaker drivers worn on or around the head over a user's ears. Headphones are electroacoustic transducers, which convert an electrical signal to a corresponding sound. Headphones enable a single user to listen to an audio source privately, in contrast to a loudspeaker, which emits sound into the open air for anyone nearby to hear. Headphones are also known as earspeakers or earphones. Circumaural ('around the ear') and supraaural ('on the ear') headphones use a band over the top of the head to hold the speakers in place. The other type, known as earbuds or earpieces consist of individual units that plug into the user's ear canal. In the context of telecommunication, a headset is a combination of headphone and microphone. Headphones connect to a signal source such as an audio amplifier, radio, CD player, portable media player, mobile phone, video game console, or electronic musical instrument, either directly using a cord, or using wireless technology such as Bluetooth or FM radio.

Acoustically closed headphones are preferred to attenuate the outside noise as much as possible and to achieve a good audio reproduction quality due to a better signal to noise ratio in noisy environments. Closed headphones, especially "intra-aural" (in-ear) and "intra-concha" (earbud) headphones which seal the ear-canal, are likely to increase the acoustic impedance seen from the inside of the ear-canal to the outside. An increased acoustic impedance is followed by an increased sound pressure level for low frequencies inside the ear canal. In the case of self-generated sound, e.g., speaking, rubbing and buzzing noise, the sound is perceived as unnaturally amplified and uncomfortable while listening or speaking. This effect is commonly described as the occlusion effect.

1

As shown in FIG. 1, the sound pressure at the eardrum is a summation of the sound generated by the larynx 101 transmitted through the mouth 104 to the beginning of the ear canal 103 and the transmitted body-conducted sound. FIG. 1 shows the open ear scenario (reference scenario) in which no occlusion effect occurs. Hme represents the transfer path from the mouth 104 to the inner ear 102 through the air, and Hb represents the transfer path from the larynx to the inner ear 102 through bones.

If the ear canal is closed with headphones, as schematically shown in FIG. 2, the sound transmitted through the air (Hme) is damped with the isolation curve of the headphone (105). The signal path through the body (Hb) remains unchanged. Sealing the ear canal with the headphone has the effect of amplifying the sound pressure in front of the ear drum in low frequencies. This causes the occlusion effect.

In FIG. 3, the occlusion effect is illustrated by comparing two sound pressure level spectra measured inside the ear canal. The dashed curve shows the sound pressure for the unoccluded open ear. The solid curve shows the sound pressure level inside the ear canal of the same subject wearing circumaural (over-ear) headphones. It can be seen that the sound pressure level is increased in the frequency range between approximately 60 Hz and 400 Hz.

"Naturalness" is one of the important perceptual attributes for sound reproduction over headphones. Naturalness is defined as the feeling of the user to be fully immersed in the original environment. In the case of a "listening only" scenario, this is a binaural recording at the entrance of the ear canal which is played back (ambient sound). From the moment the user starts to speak, the reproduction of ambient sounds is less important and the immersion is attenuated. In the scenario of a user who is speaking or participating in a teleconference, the ambient sound is less important. Therefore, it is more important to ensure that the perception of the own voice when the user wears a headset is as close as to the perception without a headset. However, the naturalness is affected by wearing acoustically closed headphones, especially the in-ear headphones, since such headphones have a strong occlusion effect.

Summary of the Invention

The main technical field of the present invention is binaural audio reproduction over headphones. It is an object of the invention to reduce the occlusion effect for in-ear or earbud head-
5   phones by capturing user's own voice with the in-line microphone of a headset, and embodiments of the invention also could be used for over-ear or on-ear headsets. The headset is processed with an anti-occlusion algorithm to create a natural sound pressure distribution inside the ear canal.

10   The foregoing and other objects are achieved by the subject matter of the independent claims. Further implementation forms are apparent from the dependent claims, the description and the figures.

A first aspect of the invention provides a method for processing an audio signal, the method
15   comprising: processing the audio signal according to a pair of mouth to ear transfer functions, to obtain a processed audio signal; filtering the processed audio signal, using a pair of equalization filters, to obtain a filtered audio signal, wherein a parameter of the equalization filter depends on an acoustic impedance of a headphone; and outputting the filtered audio signal to the headphone.
20

According to the audio processing method in the first aspect, an occlusion effect for in-ear or earbud headphones has been reduced, and a natural sound pressure distribution inside the ear canal is created.

25   An audio signal is a representation of sound, typically using a level of electrical voltage for analog signals, and a series of binary numbers for digital signals. Audio signals have frequencies in the audio frequency range of roughly 20 to 20,000 Hz, which corresponds to the upper and lower limits of human hearing). Audio signals may be synthesized directly, or may originate at a transducer such as a microphone, musical instrument pickup, phonograph cartridge, or
30   tape head. Loudspeakers or headphones convert an electrical audio signal back into sound.
In an example, an audio signal may be obtained by a receiver. For example, the receiver may obtain the audio signal from another device or another system via a wired or wireless communication channel.

In another example, an audio signal may be obtained using a microphone and a processor. The microphone is used to record information obtained from a sound source, and the processor is used to process these information recorded by the microphones, to obtain the audio signal.

5   In one implementation form of the first aspect, the mouth to ear transfer function describes a transfer function from the mouth to the eardrums.

In one implementation form of the first aspect, the mouth to ear transfer function is obtained using a head and torso simulator; or the mouth to ear transfer function is obtained using a real

10   person.

In an example, a head and torso simulator equipped with mouth and ear simulators provides an approach to the measurement of HmeTFs. The transfer functions or impulse response from an input signal (fed to the loudspeaker of the mouth simulator) to the output signals (from the ear

15   microphones) are measured.

In another example, a transfer function can be measured from a microphone or a speaker near the mouth to the ear microphones. Compared with the above example which refers to a head and torso simulator, using a real person has the advantage of removing the response of the

20   mouth simulator from the measurement, and is also well-suited to simulation –as a talking subject can have a microphone positioned similarly near their mouth as part of the simulation system.

Equalization is the process of adjusting the balance between frequency components within an

25   electronic signal. In sound recording and reproduction, equalization is the process commonly used to alter the frequency response of an audio system using linear filters or other type filters. The circuit or equipment used to achieve equalization is called an equalization filter or an equalizer. These devices strengthen (boost) or weaken (cut) the energy of specific frequency bands or "frequency ranges".

30

Common equalizers or filters in music production are parametric, semi-parametric, graphic, peak, and program equalizers or filters. Graphic equalizers or filters are often included in con-

4

sumer audio equipment and software which plays music on home computers. Parametric equalizers or filters require more expertise than graphic equalizers, and they can provide more specific compensation or alteration around a chosen frequency. This may be used in order to remove unwanted resonances or boost certain frequencies.

5

Acoustic impedance is the ratio of acoustic pressure to flow. In an example, the acoustic impedance according to the standard ISO-10534-2 is defined as the "ratio of the complex sound pressure p(0) to the normal component of the complex sound particle velocity v(0) at an individual frequency in the reference plane". The reference plane is the cross-section of the impedance tube for which the impedance Z (or the reflection factor r, or the admittance G) are determined and is a surface of a test object. The reference plane is assumed to be at x= (in our context, this is the end of the tube, where the test object starts). Therefore, p(0) describes the complex sound pressure at the end of the tube and v(0) describes the complex particle velocity at the end of the tube. Complex sound pressure p and complex sound pressure v denote the Fourier transform of these parameters in the time domain.

In an example, for a linear time-invariant system, the relationship between the acoustic pressure applied to the system and the resulting acoustic volume flow rate through a surface perpendicular to the direction of that pressure at its point of application is given by

$$p(t) = [R * Q](t),$$

or equivalently by

$$Q(t) = [G * p](t),$$

where

- p is the acoustic pressure;
- Q is the acoustic volume flow rate;
- $*$ is the convolution operator;
- R is the acoustic resistance in the *time domain*;
- $G = R^{-1}$ is the acoustic conductance in the *time domain* (R^{-1} is the convolution inverse of R).

Acoustic impedance, denoted Z, is the Laplace transform, or the Fourier transform, or the analytic representation of *time domain* acoustic resistance:

5

$$Z(s) \overset{\text{def}}{=} \mathcal{L}[R](s) = \frac{\mathcal{L}[p](s)}{\mathcal{L}[Q](s)},$$

$$Z(\omega) \overset{\text{def}}{=} \mathcal{F}[R](\omega) = \frac{\mathcal{F}[p](\omega)}{\mathcal{F}[Q](\omega)},$$

$$Z(t) \overset{\text{def}}{=} R_a(t) = \frac{1}{2}\left[p_a * \left(Q^{-1}\right)_a \right](t),$$

where

- $\mathcal{L}$ is the Laplace transform operator;
- $\mathcal{F}$ is the Fourier transform operator;
- subscript "a" is the analytic representation operator;
- Q^{-1} is the convolution inverse of Q.

In one implementation form of the first aspect, the acoustic impedance of the headphone is measured based on an acoustic impedance tube. The acoustic impedance tube may have a measurable frequency range from 20 Hz to 2 kHz, for example.

In one implementation form of the first aspect, the parameter of the equalization filter is a gain factor of the equalization filter, the gain factor of the equalization filter is proportional to the inverse of the acoustic impedance of the headphone.

In an example, a gain factor or a shape (g) of an equalization filter is proportional to the inverse of Z_{HP}.

$$g = \alpha \, \frac{1}{Z_{HP}},$$

where α is the scaling factor (proportional coefficient), which can either be selected by the user or determined during measurements of different headphones.

In one implementation form of the first aspect, the pair of equalization filters is selected based on a headphone type of the headphone.

In an example, the equalization filter is pre-designed based on the acoustic impedance of the headphone. Therefore, information of the headphone used is required. Select the headphone type can be done either manually or automatically. For example, the headphone type can be

selected by the user manually based on the headphone categories (e.g., over-ear headphone, on-ear headphone) or the headphone model (e.g., HUAWEI Earbud). The headphone type can also be selected automatically detected by the information provided by the USB type-C. For each headphone, the equalization filter is then chosen based on the headphone's acoustic impedance, as mentioned above. For each category, we can design the filter based on an averaged acoustic impedance or use a representative equalization filter for each category.

In one implementation form of the first aspect, the headphone type of the headphone is obtained based on a Universal Serial Bus, USB, Type-C information.

A second aspect of the invention provides an apparatus for processing a stereo signal, wherein the apparatus comprises processing circuitry configured to: process the audio signal according to a pair of mouth to ear transfer functions, to obtain a processed audio signal; filter the processed audio signal, using a pair of equalization filters, to obtain a filtered audio signal, wherein a parameter of the equalization filter depends on an acoustic impedance of a headphone; and output the filtered audio signal to the headphone.

The processing circuitry may comprise hardware and software. The hardware may comprise analog or digital circuitry, or both analog and digital circuitry. In one embodiment, the processing circuitry comprises one or more processors and a non-volatile memory connected to the one or more processors. The non-volatile memory may carry executable program code which, when executed by the one or more processors, causes the apparatus to perform the operations or methods described herein.

In one implementation form of the second aspect, the mouth to ear transfer function describes a transfer function from the mouth to the eardrums.

In one implementation form of the second aspect, the acoustic impedance of the headphone is measured based on an acoustic impedance tube, the acoustic impedance tube has a measurable frequency range from 20 Hz to 2 kHz.

In one implementation form of the second aspect, wherein the parameter of the equalization filter is a gain factor of the equalization filter, the gain factor of the equalization filter is proportional to the inverse of the acoustic impedance of the headphone.

5　In one implementation form of the second aspect, wherein the pair of equalization filters is selected based on a headphone type of the headphone.

In one implementation form of the second aspect, the headphone type of the headphone is obtained based on a Universal Serial Bus, USB, Type-C information.

10

The filters described in this disclosure may be implemented in hardware or in software or in a combination of hardware and software.

A third aspect of the invention relates to a computer-readable storage medium storing program

15　code. The program code comprises instructions for carrying out the method of the first aspect or one of its implementations.

The invention can be implemented in hardware and/or software.

20　　　　　　　　　　　　　**Brief Description of the Drawings**

To illustrate the technical features of embodiments of the present invention more clearly, the accompanying drawings provided for describing the embodiments are introduced briefly in the following. The accompanying drawings in the following description are merely some embodi-

25　ments of the present invention, but modifications on these embodiments are possible without departing from the scope of the present invention as defined in the claims.

FIG. 1 shows an example about open ear scenario (reference scenario) in which no occlusion effect occurs.

30　FIG. 2 shows an example of an ear scenario (reference scenario) in which occlusion effect occurs.

FIG. 3 shows an example about the occlusion effect by comparing two sound pressure level spectra measured inside the ear canal.

FIG. 4 shows a schematic diagram of a method for reducing the occlusion effect according to an embodiment,

FIG. 5 shows an example of measurement of a mouth to ear transfer function.

FIG. 6 shows a schematic diagram of measure an acoustic impedance of a headphone by using an acoustic impedance tube according to an embodiment;

FIG. 7 shows an example of an acoustic impedance of open headphone and an acoustic impedance of closed headphone;

FIG. 8 shows an example of acoustic impedances for an in-ear headphone and an ear-bud headphone;

FIG. 9 shows an example of frequency curve for an equalization filter;

FIG. 10 shows a signal processing chart of a method of using a telephone with a headset in a quiet environment according to an embodiment;

FIG. 11 shows an example of a high-pass shelving filter according to an embodiment;

FIG. 12 shows a signal processing chart of a method of using a telephone with a headset in a noisy environment according to an embodiment;

FIG. 13 shows a signal processing chart of a method for processing an audio signal according to an embodiment;

FIG. 14 shows a schematic diagram illustrating a device for processing an audio signal according to an embodiment.

In the figures, identical reference signs will be used for identical or functionally equivalent features.

Detailed Description of the Embodiments

5

In the following description, reference is made to the accompanying drawings, which form part of the disclosure, and in which are shown, by way of illustration, specific aspects in which the invention may be placed. It will be appreciated that the invention may be placed in other aspects and that structural or logical changes may be made without departing from the scope of the

10   invention. The following detailed description, therefore, is not to be taken in a limiting sense, as the scope of the invention is defined by the appended claims.

For instance, it will be appreciated that a disclosure in connection with a described method will generally also hold true for a corresponding device or system configured to perform the method

15   and vice versa. For example, if a specific method step is described, a corresponding device may include a unit to perform the described method step, even if such unit is not explicitly described or illustrated in the figures.

Moreover, in the following detailed description as well as in the claims, embodiments with

20   functional blocks or processing units are described, which are connected with each other or exchange signals. It will be appreciated that the invention also covers embodiments which include additional functional blocks or processing units, such as pre- or post-filtering and/or pre- or post-amplification units, that are arranged between the functional blocks or processing units of the embodiments described below.

25

Finally, it is understood that the features of the various exemplary aspects described herein may be combined with each other, unless specifically noted otherwise.

A channel is a pathway for passing on information, in this context sound information. Physically, it might, for example, be a tube you speak down, or a wire from a microphone to an

30   earphone, or connections between electronic components inside an amplifier or a computer.

A track is a physical home for the contents of a channel when recorded on magnetic tape. There can be as many parallel tracks as technology allows, but for everyday purposes there are 1, 2 or

4. Two tracks can be used for two independent mono signals in one or both playing directions, or a stereo signal in one direction. Four tracks (such as a cassette recorder) are organized to work pairwise for a stereo signal in each direction; a mono signal is recorded on one track (same track as the left stereo channel) or on both simultaneously (depending on the tape recorder or on how the mono signal source is connected to the recorder).

A mono sound signal does not contain any directional information. In an example, there may be several loudspeakers along a railway platform and hundreds around an airport, but the signal remains mono. Directional information cannot be generated simply by sending a mono signal to two "stereo" channels. However, an illusion of direction can be conjured from a mono signal by panning it from channel to channel.

A stereo sound signal may contain synchronized directional information from the left and right aural fields. Consequently it requires at least two channels, one for the left field and one for the right field. The left channel is fed by a mono microphone pointing at the left field and the right channel by a second mono microphone pointing at the right field (you will also find stereo microphones that have the two directional mono microphones built into one piece). In an example, *Quadraphonic stereo* uses four channels, *surround stereo* has at least additional channels for anterior and posterior directions apart from left and right. Public and home cinema stereo systems can have even more channels, dividing the sound fields into narrower sectors.

Stereophonic sound or, more commonly, stereo, is a method of sound reproduction that creates an illusion of multi-directional audible perspective. This is usually achieved by using two or more independent audio channels through a configuration of two or more loudspeakers (or stereo headphones) in such a way as to create the impression of sound heard from various directions, as in natural hearing.

In one embodiment of the present invention, the object of the audio signal processing method or audio signal processing apparatus is to improve the naturalness and to reduce the occlusion effect when using in-ear headphones, and to counteract the occlusion effect and to provide a sound pressure that will be perceived as natural. In an example, the user's voice is captured by the in-line microphone and convolved 402 with a pair of mouth to ear transfer function

(HmeTF) 401 for left/right ear form a recorded or a database, respectively (FIG. 4). The resulting signal is filtered (k) with an equalization filter (anti-occlusion filter) 403 which is designed based on the acoustic impedance of the used headphone.

A head-related transfer function (HRTF) is a response that characterizes how an ear receives a sound from a point in space. As sound strikes the listener, the size and shape of the head, ears, ear canal, density of the head, size and shape of nasal and oral cavities, all transform the sound and affect how it is perceived, boosting some frequencies and attenuating others. Generally speaking, the HRTF boosts frequencies from 2–5 kHz with a primary resonance of +17 dB at 2,700 Hz.

A pair of HRTFs for two ears can be used to synthesize a binaural sound that seems to come from a particular point in space. It is a transfer function, describing how sound from a specific point will arrive at the ear (generally at the outer end of the auditory canal). HRTFs for left and right ear describe the filtering of sound by the sound propagation paths from the source to the left and right ears, respectively. The HRTF can also be described as the modifications to a sound from a direction in free air to the sound as it arrives at the eardrum.

The mouth to ear transfer function (HmeTF) describes the transfer function from the mouth to the eardrums. HmeTF can be measured non-individually by using a dummy head (head-torso with mouth simulator), or HmeTF can be measured individually by placing a smartphone or microphone close to the mouth of a user and reproducing a measurement signal. The measurement signal is acquired by microphones placed near the entrance of the blocked ear canal (120). The measurement signal can be a noise signal. FIG. 5 shows an example of a measurement of an individual HmeTF. If a non-individual HmeTF is used, it can be measured once and provided to many users. If an individual HmeTF is required, it needs to be measured once for each user.

In an example, a HmeTF measurement can be made of a real room environment from the mouth to the ears of the same head. For simulation, a talker's voice is convolved in real-time with the HmeTF, so that the talker can hear the sound of his or her own voice in the simulated room environment. We show by example how HmeTF measurements can be made using human subjects (by measuring the transfer function of speech) or by a head and torso simulator.

In an example, a HmeTF is measured using a head and torso simulator (HATS). The mouth simulator directivity of the HATS is similar to the mean long term directivity of conversational speech from humans, except in the high frequency range. The HATS' standard mouth microphone position (known as the 'mouth reference point') is 25 mm away from the 'center of lip'

5   (which in turn is 6 mm in front of the face surface). A microphone is used at the mouth reference point. Rather than using the inbuilt microphones of the HATS (which are at the acoustic equivalent to eardrum position), some microphones that are positioned near the entrance of the ear canals are used. One reason is that a microphone setup similar to the one of the HATS is used on a real person. The microphone setup on the real person comprises microphones which may

10   be similar or identical to the microphones of the HATS microphones and which are placed at positions equivalent to those of the HATS. Another reason is that it is desirable to avoid measuring with ear canal resonance, as the strong resonant peaks would need to be inverted in the simulation, which would introduce noise and perhaps latency.

15   In another example, the measurement about the HmeTF is made by sending a swept sinusoid test signal to the mouth loudspeaker, the sound of which was recorded at the mouth and ear microphones. The sweep ranged between 50 Hz – 15 kHz, with a constant sweep rate on the logarithmic frequency scale over a period of 15 s. A signal suitable for deconvolving the impulse response from the sweep was sent directly to the recording device, along with the three

20   microphone signals. This yielded the impulse response (IR) from the signal generator to a microphone, and the transfer function is obtained from the mouth microphone to ear microphones by dividing the latter by the former in the frequency domain. The procedure for this is, first, to take the Fourier transform of the direct sound from the mouth microphone impulse response, zero-padded to be twice the length of the desired impulse response. The direct sound is identi-

25   fied by the maximum absolute value peak of the mouth microphone IR, and data from -2 to +2 ms around this is used, with a Tukey window function applied (50% of the window is fade-in and fade-out using half periods of a raised cosine, and the central 50% has a constant coefficient of 1).

30   In another example, a Fourier transform window length is used for the ear microphone impulse responses, with the second half of the window zero-padded. The transfer function is obtained by dividing the cross-spectrum (conjugate of mouth IR multiplied by the ear IR) by the auto-spectrum of the mouth microphone's direct sound. Before returning to the time domain, a band-

pass filter is applied to the transfer function to be within 100 Hz – 10 kHz to avoid signal-to-noise ratio problems at the extremes of the spectrum (this is done by multiplying the spectrum components outside this range by coefficients approaching zero). After applying an inverse Fourier transform, the impulse response is truncated (discarding the latter half). The resulting

5   IR for each ear is multiplied by the respective ratio of mouth-to-ear rms values of microphone calibration signals (sound pressure level of 94 dB) to compensate for differences in gain between channels of the recording system.

In another example, HmeTFs can be measured using a real person and using a microphone

10   arrangement similar or identical to the one used in a HATS. The sound source could simply be speech, although other possibilities exist. The transfer function is calculated between a microphone near the mouth to each of the ear microphones. This approach was taken in measuring the transfer function from mouth to ear (without room reflections), and it can be used for measuring room reflections too. Advantages of using such a technique (compared to using the

15   HATS) may include matching the individual long term speech directivity of the person; matching the head related transfer functions of the person's ears; and that the measurement system only requires minimal equipment.

In an example, the formula of the HmeTF depends on how it is measured, generally it is the

20   ratio between the complex sound signal at the ear and at the mouth, $HmeTF = p\_ear / p\_mouth$.

In another example, the HmeTF is measured using a real person and a smartphone. The microphone setup can be similar to the other examples and the smartphone has to be positioned near to the mouth. The smartphone acts as a sound source and as the reference microphone. The

25   transfer function is calculated between the smartphone microphone (reference microphone) and the ear microphones. The advantages of this method is the increased bandwidth of the sound source compared with the speech of the real person.

Parameters of the equalization filter are based on the acoustic impedance of the headphone. The

30   acoustic impedance of the headphone in low frequency is highly correlated with the perceived occlusion effect, i.e., high acoustic impedance corresponds to high occlusion effect caused by the headphone. The acoustic impedance of the headphone can be measured using a customized acoustic impedance tube, e.g. an acoustic impedence tube built in accordance with ISO-10534-

2.The measurement tube may be built to fit the geometries of a human ear canal, e.g., the inner diameter of the tube should be approx. 8 mm, and a frequency range should be between at least 60 Hz and 2 kHz. As shown in FIG. 6, the acoustic impedances of 1) the artificial ear with headphone ($\underline{Z}_{OE}^{Hp}$) and 2) the artificial ear without headphone ($\underline{Z}_{OE}$) are measured. Then the acoustic impedance of the headphone (Z_{HP}) can be determined by calculating the ratio between the $\underline{Z}_{OE}^{Hp}$ and $\underline{Z}_{OE}$:

$$Z_{HP} = \frac{|\underline{Z}|_{OE}^{Hp}}{|\underline{Z}|_{OE}}.$$

In another example, the acoustic impedance of the headphone (Z_{HP}) may be determined by calculating the difference between the $\underline{Z}_{OE}^{Hp}$ and $\underline{Z}_{OE}$:

$$Z_{HP} = \underline{Z}_{OE}^{Hp} - \underline{Z}_{OE}.$$

FIG. 7 shows an example of the acoustic impedance of open and closed headphones. The dashed line shows the acoustic impedance of an open headphone. The perceived occlusion effect for the open headphone is very low. The solid line shows the acoustic impedance of a closed ear-phone. The increased impedance in the low frequency range up to 1.5 kHz boosts the low frequency sound level, which corresponds to a high perceived occlusion.

The curves 110, 111 in FIG.8 show exemplary results of acoustic impedances for in-ear head-phones 110 and ear-bud headphones 111. The impedance of ear-bud headphones 111 is lower than the impedance of in-ear headphones 110 up to a frequency of about 1 kHz.

The gain factor/shape (g) of an equalization filter is proportional to the inverse of Z_{HP}.

$$g = \alpha \, \frac{1}{Z_{HP}},$$

where α is the scaling factor (proportional coefficient), which can either be selected by the user or determined during a lot of measurement of different headphones. FIG. 9 shows the exemplary target frequency curve for the equalization filter to reduce the occlusion effect and to improve naturalness of user's own voice. 112 shows the target response curve for the in-ear headphone, 113 the target response curve for the ear bud.

FIG. 13 shows a schematic diagram of a method for processing an audio signal according to an embodiment. The method comprises:

S21: processing the audio signal according to a pair of mouth to ear transfer functions.

S22: filtering the processed audio signal, using a pair of equalization filters.

S23: outputting the filtered audio signal to the headphone.

5   Embodiment 1, telephone with headset (in-ear headphone or earbuds with in-line microphone) in a quiet environment.

FIG. 10 shows a block diagram of this embodiment. User's own voice (air-transmitted) is captured using an in-line microphone of the headphone used. The captured speech signal 13 is

10   filtered through a pair of mouth-to-ear transfer functions (HmeTFs), which can be individually or non-individually determined 14 before. The filtered speech signals are then further filtered through a pair of anti-occlusion hear-through equalization filters to enhance the high pass component of user's own voice. The filtered signals are then played back using headphones to the user and the naturalness while the user is speaking is enhanced.

15

The anti-occlusion hear-through equalization filter 12 is pre-designed based on the acoustic impedance of the headphone. Therefore, information of the headphone used is required. It can be done either manually or automatically. For example, the headphone can be selected 11 by the user manually based on the headphone categories (e.g., over-ear headphone, on-ear head-

20   phone) or the headphone model (e.g., HUAWEI Earbud). It can also be automatically detected by the information provided by the USB type-C. For each headphone, the anti-occlusion hear-through equalization filter is then chosen based on its acoustic impedance, as mentioned above. For each category, we can design the filter based on an averaged acoustic impedance or use a representative equalization filter for each category.

25

The shape of the filter should be proportional to the inverse of the acoustic impedance ($0-Z_{HP}$ in dB). For the design of the anti-occlusion hear-through equalization filter, almost every low order Infinite Impulse Response, IIR, filter or finite impulse response, FIR, filter is suitable (low latency is needed).

30

FIG. 11 shows an example in which a high-pass shelving filter (FIR-filter) is used for the design of an anti-occlusion hear-through equalization filter in one implementation. Also other filters, e.g. an implementation with a Chebyshev-II IIR-filter, can be used.

The filter can be designed in two steps:

1) The stopband attenuation can be determined by averaged acoustic impedance from low (60 Hz) to the cut-off frequency as a starting point. Then the cut-off frequency can be determined by the first zero crossing of the frequency dependent acoustic impedance, seen from the low to the high frequency.

2) Iterating the stopband attenuation and the cut-off frequency by minimizing the error between the inverse of the acoustic impedance curve (target) and the designed frequency response (e.g., using machine learning).

For example, the cut-off frequency is 3.5 kHz of an in-box Earbuds, and the stopband attenuation is 16 dB. The pre-designed filters can be stored in the cloud, in an online database provided to user or in the smartphone, for example.

Embodiment 2, telephone with headset (in-ear headphone or earbuds with in-line microphone) in a noisy environment.

As an example, a user is making a teleconference with a headset in a noisy room, e.g., a restaurant or an airport. The user's own voice captured by the in-line microphone is combined with the environment noise, and this may decrease the naturalness perception. In addition, the user does not want the remote user to hear the environment noise as this may reduce the speech intelligibility.

Therefore, in the case of noisy environments, the captured user's voice is first de-composed into direct sound and ambient sound. The ambient sound is discarded. The extracted direct sound is filtered through a pair of HmeTFs and is further filtered through a pair of anti-occlusion hear-through equalization filters to simulate the direct sound part. The measured or synthesized late reverberation part is added to the direct part to simulate the quite environment but with local room information. The signals are then played back using headphones to the user and the naturalness while user is speaking is enhanced. In addition, the extracted direct sound can be sent to the remote user to enhance the speech intelligibility.

In one embodiment, the binaural signals are the sum of direct sound, early reflections and late reverberation:

$$Left = d_{left}(t) + e_{left}(t) + l_{left}(t)$$
$$Right = d_{right}(t) + e_{right}(t) + l_{right}(t)$$

FIG. 14 shows a schematic diagram of a device 30 for processing an audio signal according to an embodiment. The device 30 comprises a processor 31 and a computer-readable storage medium 32 storing program code. The program code comprises instructions for carrying out embodiments of the method for processing an audio signal or one of its implementations.

Applications of embodiments of the invention include any sound reproduction system or surround sound system using multiple loudspeakers. In particular, embodiments of the presented invention can be applied to

-   TV speaker systems,
-   car entertaining systems,
-   teleconference systems, and/or
-   home cinema system,

where personal listening environments for one or multiple listeners is desirable.

The foregoing descriptions are only implementation manners of the present invention, the protection of the scope of the present invention is not limited to this. Any variations or replacements can be easily made by a person skilled in the art. The scope of protection of the present application is defined by the attached claims.

CLAIMS

1.   A method for processing an audio signal, the method comprising:

processing the audio signal according to a pair of mouth to ear transfer functions, to obtain a processed audio signal;

filtering the processed audio signal, using a pair of equalization filters, to obtain a filtered audio signal, wherein a parameter of the equalization filter depends on an acoustic impedance of a headphone; and

outputting the filtered audio signal to the headphone.

2.   The method of claim 1, wherein the mouth to ear transfer function describes a transfer function from the mouth to the eardrums.

3.   The method of claim 1 or 2, wherein the acoustic impedance of the headphone is measured based on an acoustic impedance tube, the acoustic impedance tube having a measurable frequency range from 20 Hz to 2 kHz.

4.   The method of any one of claims 1 to 3, wherein the parameter of the equalization filter is a gain factor of the equalization filter, wherein the gain factor of the equalization filter is proportional to the inverse of the acoustic impedance of the headphone.

5.   The method of any one of claims 1 to 4, wherein the pair of equalization filters is selected based on a headphone type of the headphone.

6.   The method of claim 5, wherein the headphone type of the headphone is obtained based on a Universal Serial Bus, USB, Type-C information.

7.   An apparatus for processing a stereo signal (20), wherein the apparatus (20) comprises processing circuitry (21, 22, 23, 24) configured to:

process the audio signal according to a pair of mouth to ear transfer functions, to obtain a processed audio signal;

filter the processed audio signal, using a pair of equalization filters, to obtain a filtered audio signal, wherein a parameter of the equalization filter depends on an acoustic impedance of a headphone; and

output the filtered audio signal to the headphone.

5

8.   The apparatus of claim 7, wherein the mouth to ear transfer function describes a transfer function from the mouth to the eardrums.

9.   The apparatus of claim 7 or 8, wherein the acoustic impedance of the headphone is

10   measured based on an acoustic impedance tube, the acoustic impedance tube having a measurable frequency range from 20 Hz to 2 kHz.

10.   The apparatus of any one of claims 7 to 9, wherein the parameter of the equalization filter is a gain factor of the equalization filter, the gain factor of the equalization filter being

15   proportional to the inverse of the acoustic impedance of the headphone.

11.   The apparatus of any one of claims 7 to 10, wherein the pair of equalization filters is selected based on a headphone type of the headphone.

20   12.   The apparatus of claim 11, wherein the headphone type of the headphone is obtained based on a Universal Serial Bus, USB, Type-C information.

13.   A computer-readable storage medium (32) storing program code which, when executed by a computer, causes the computer to carry out the method of any one of claims 1 to 6.

25
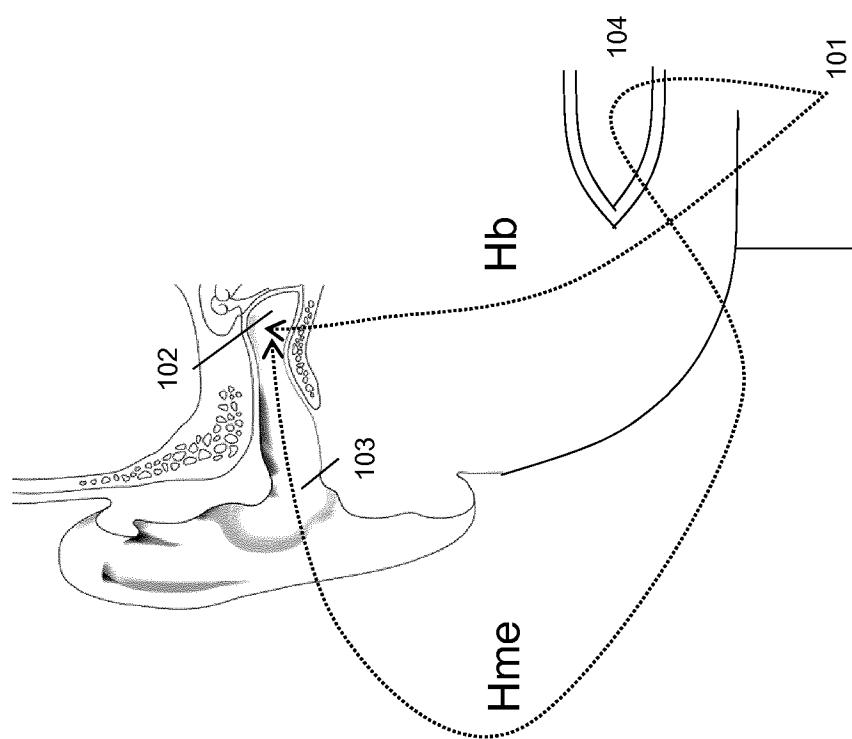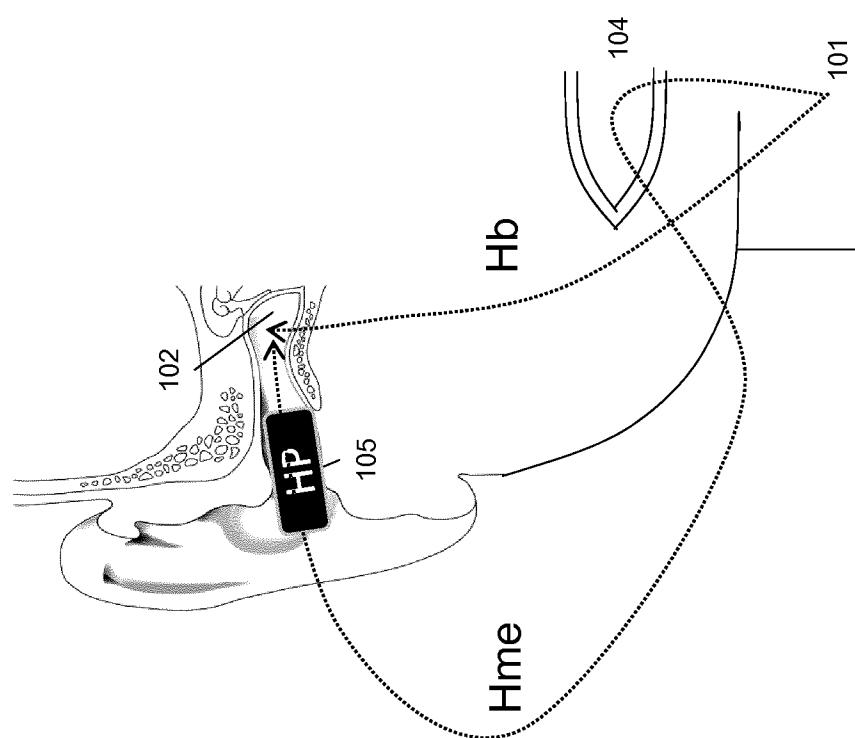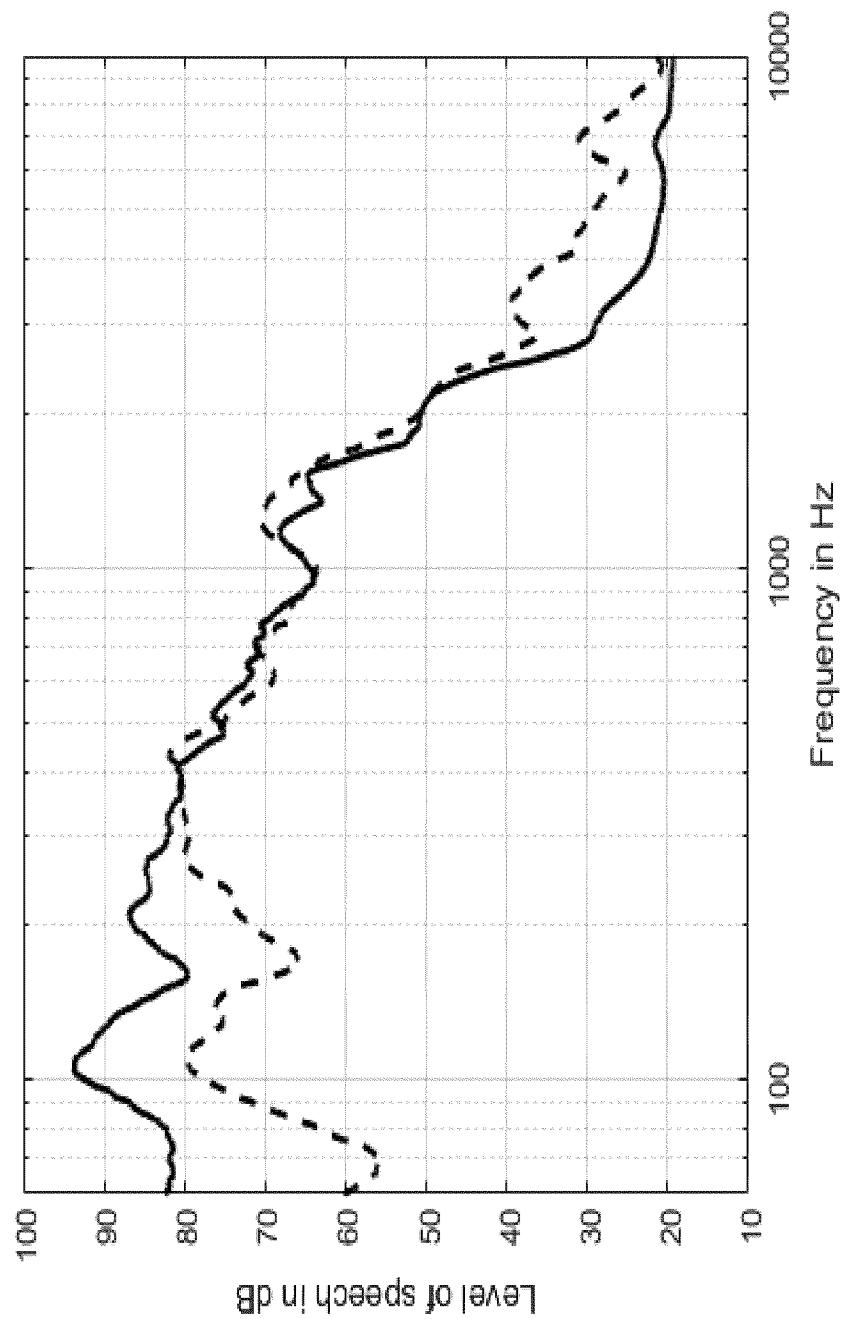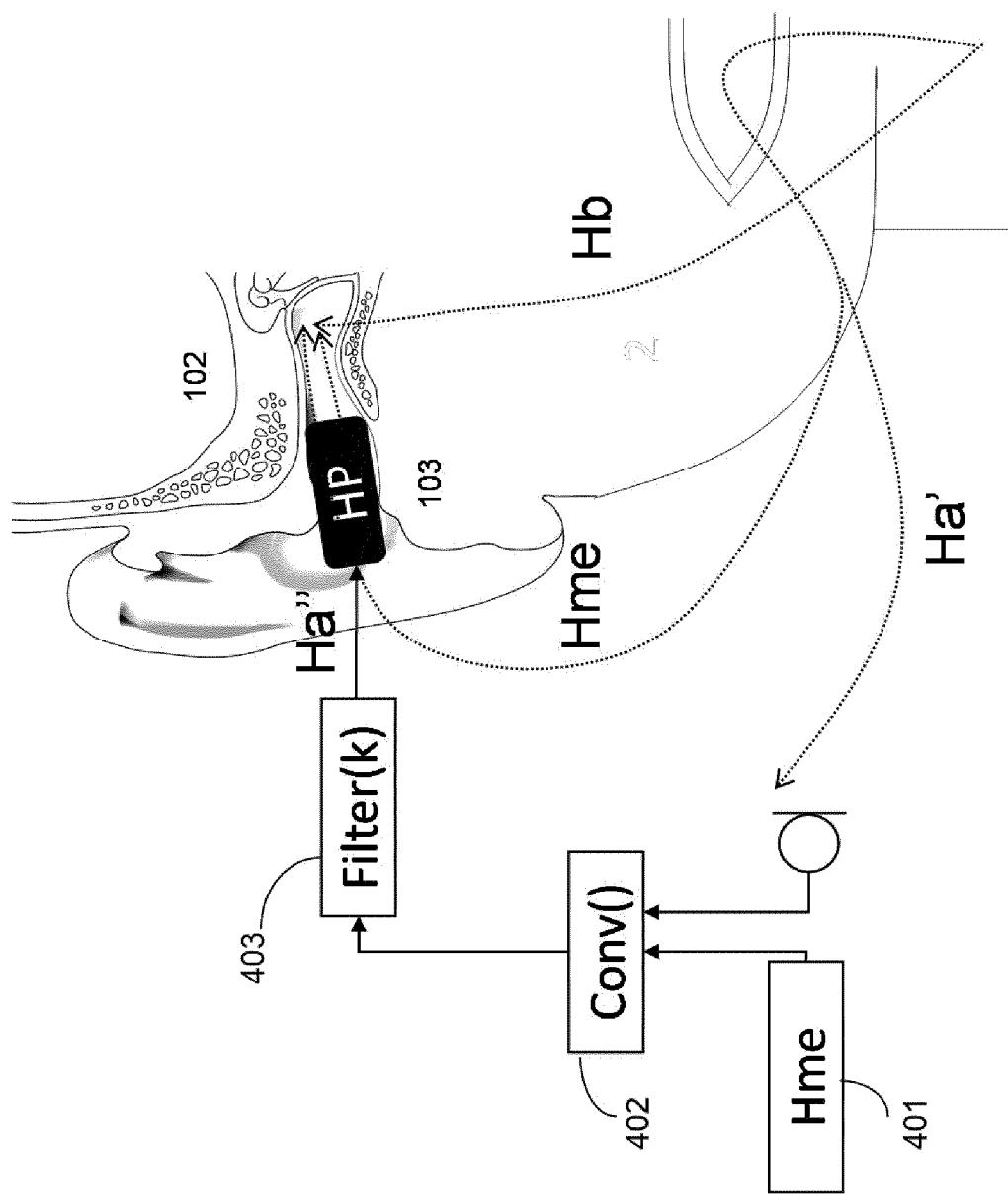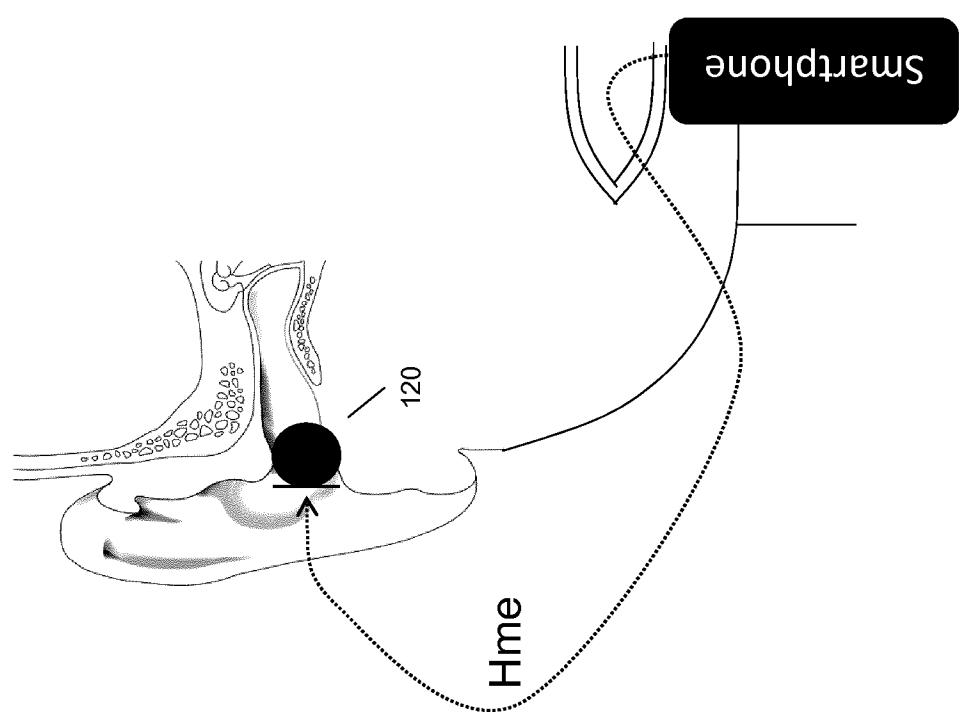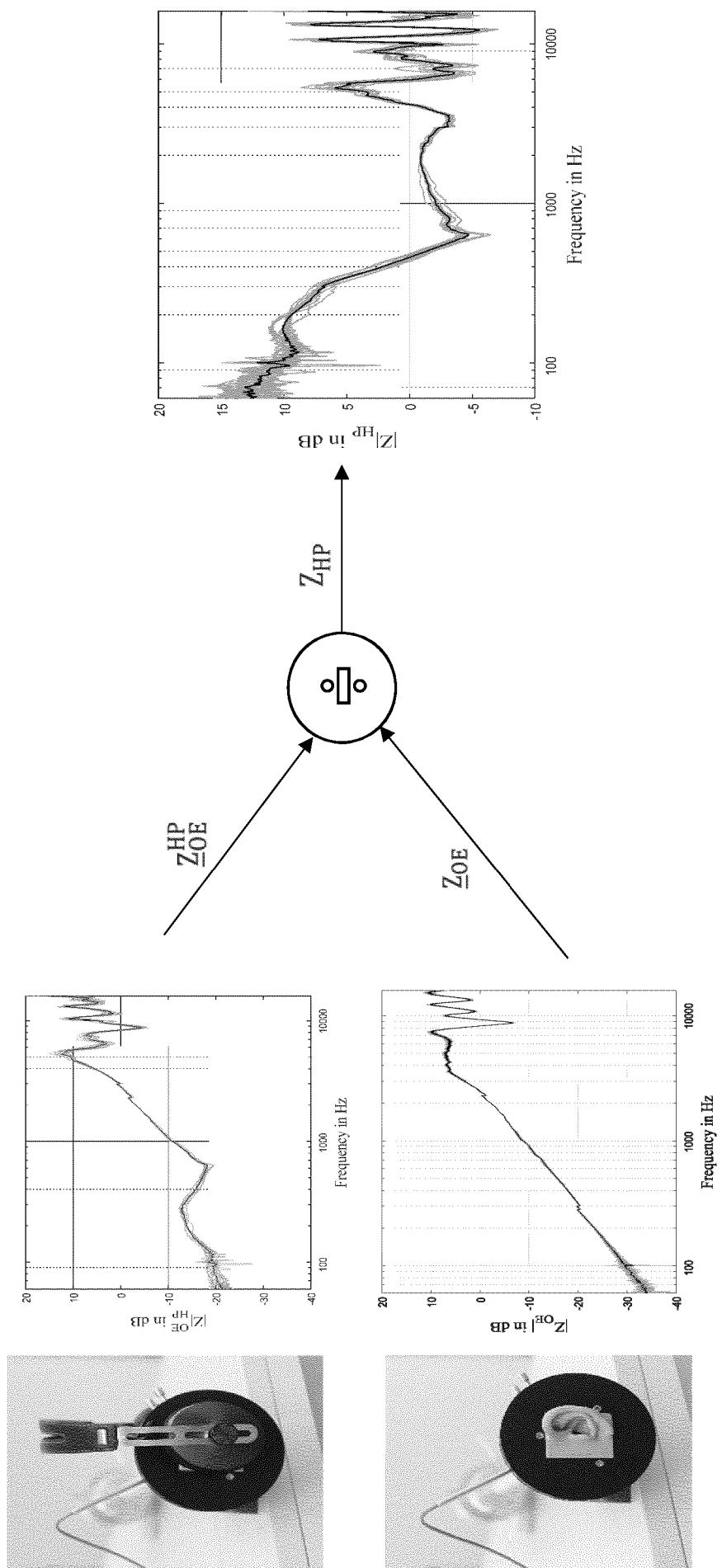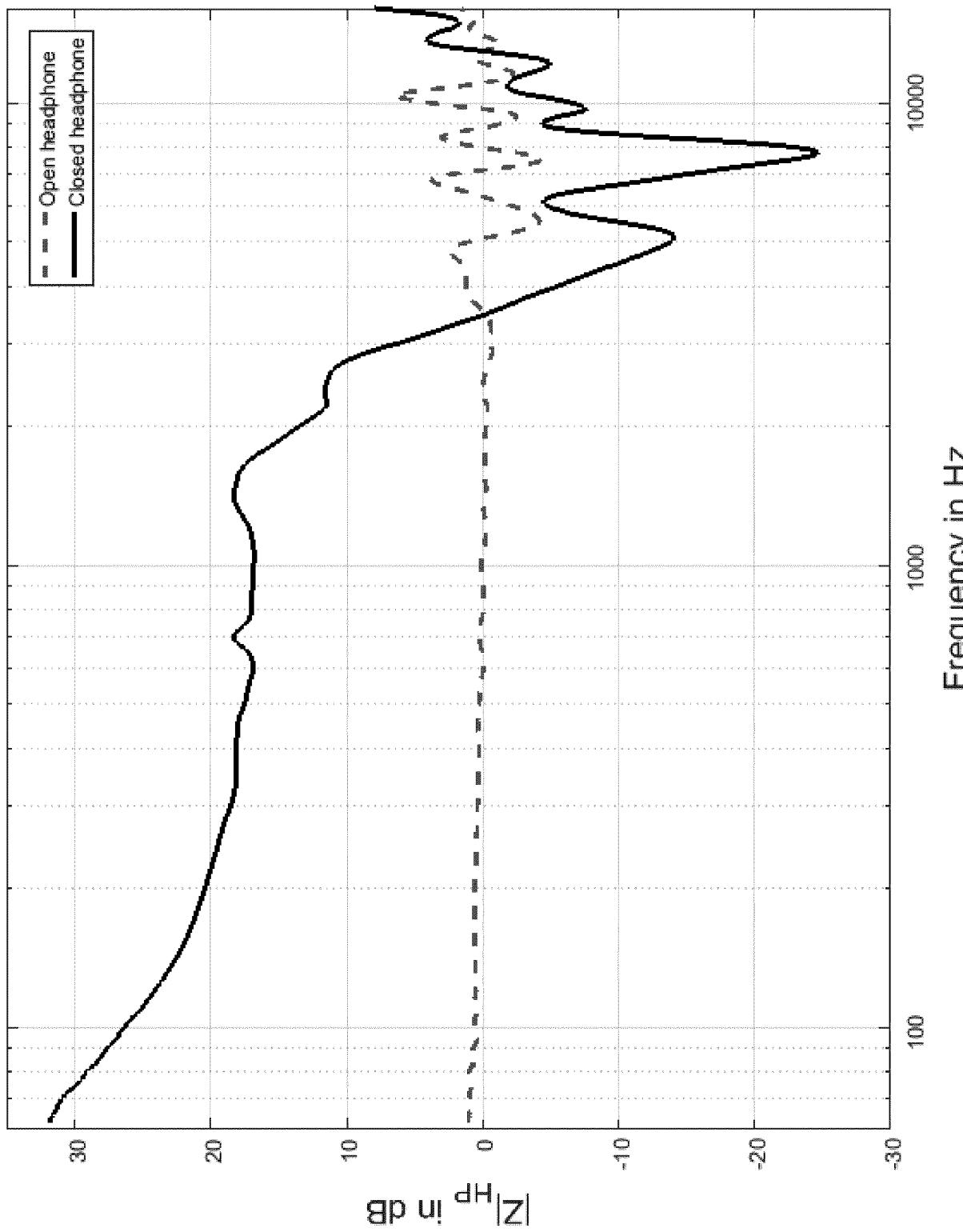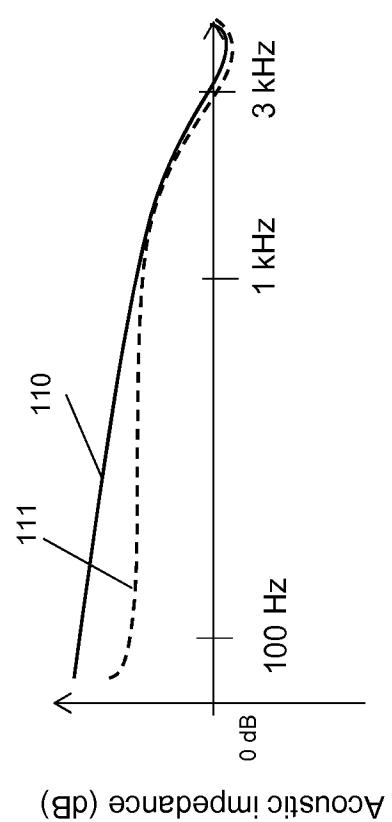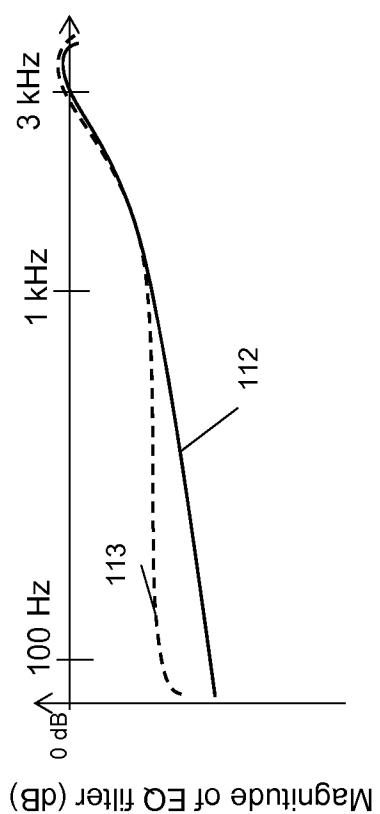
FIG. 1

FIG. 2

FIG. 3

FIG. 4

120

Hme

Smartphone

FIG. 5

FIG. 6

FIG. 7

FIG. 8

FIG. 9

FIG. 10

FIG. 11

FIG. 12

processing the audio signal according to a pair of mouth to ear transfer functions — S21

filtering the processed audio signal using a pair of equalization filters — S22

outputting the filtered audio signal to the headphone — S23

FIG. 13

Device 30

Processor 31

Memory 32
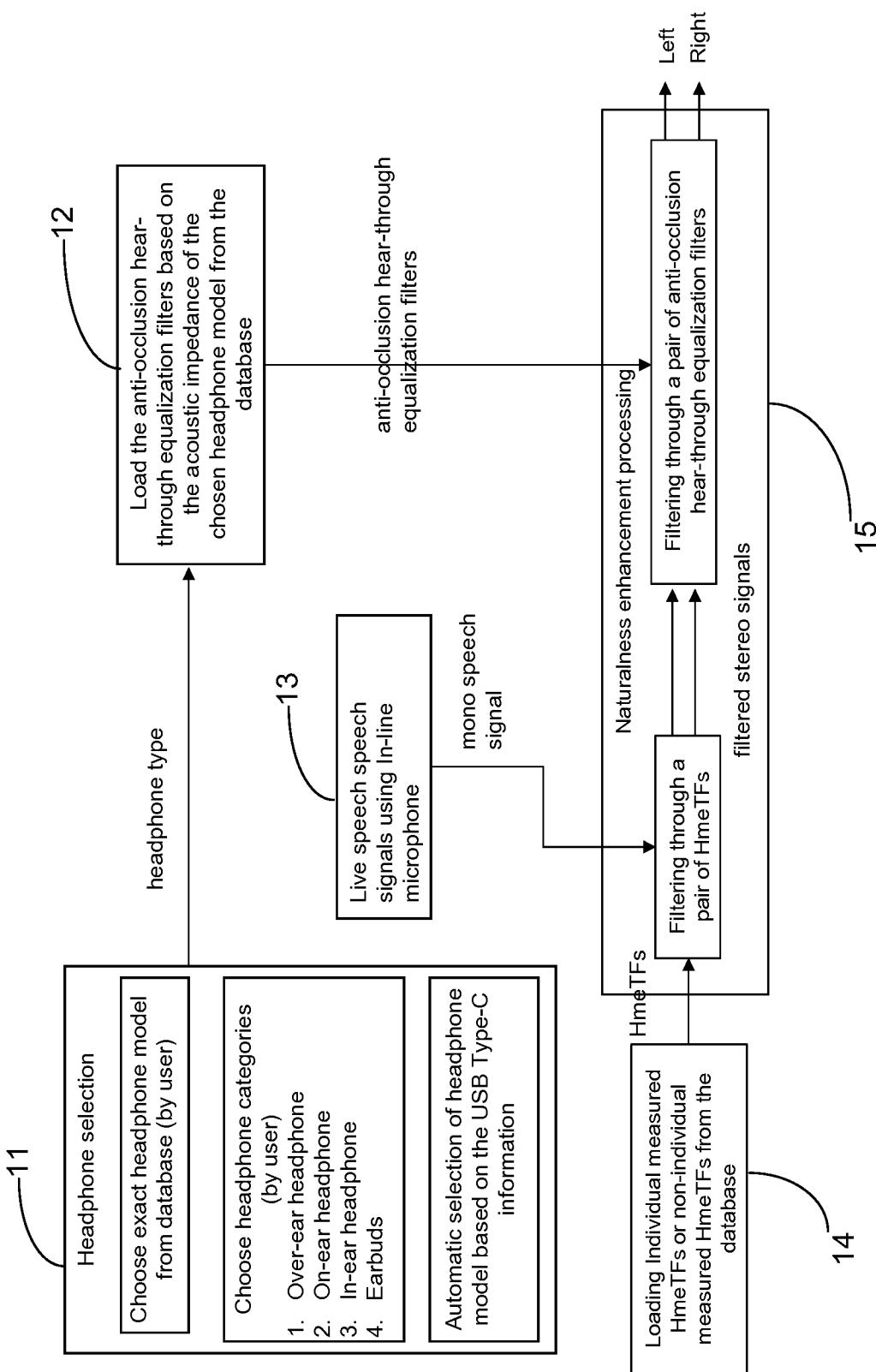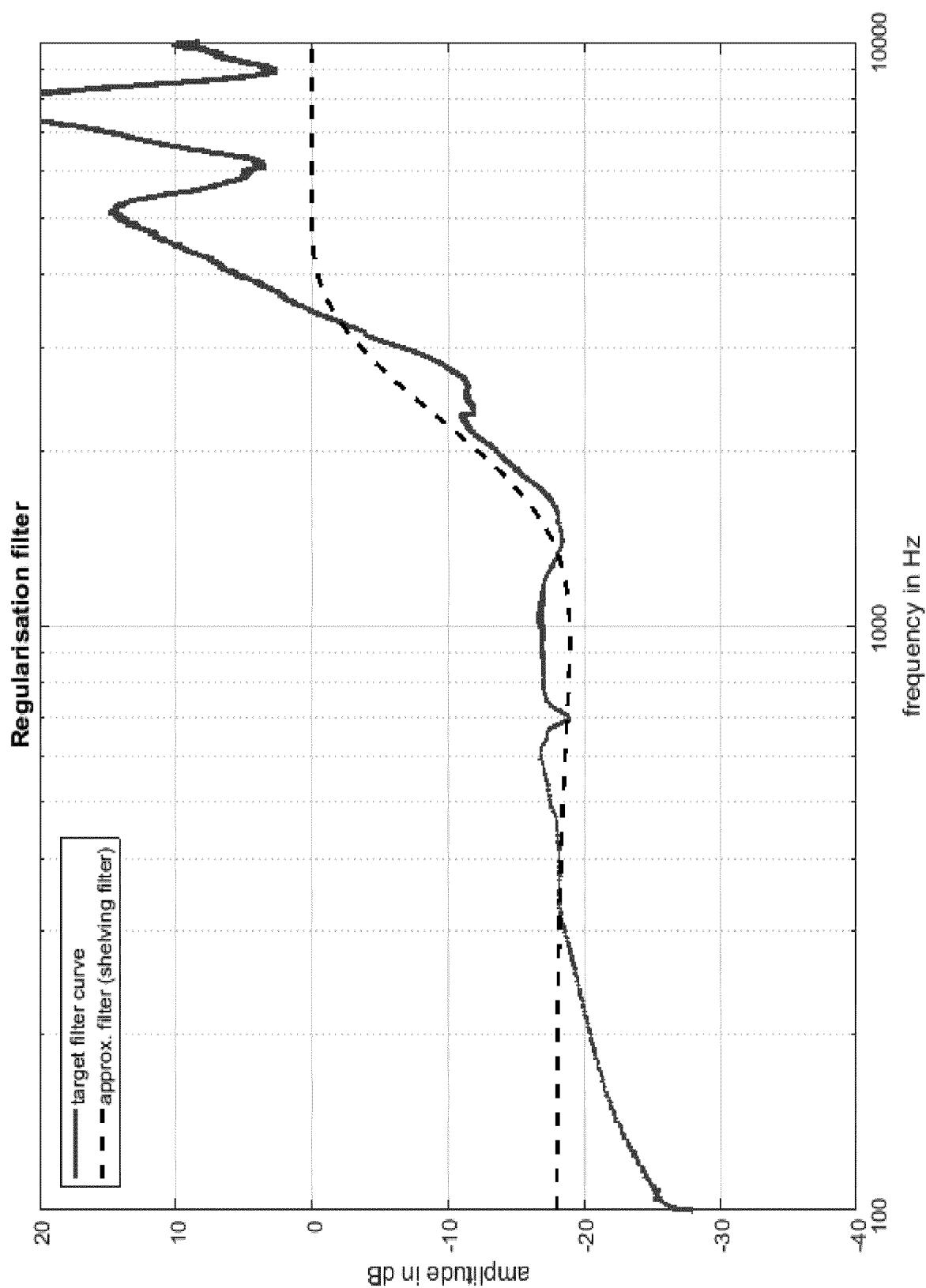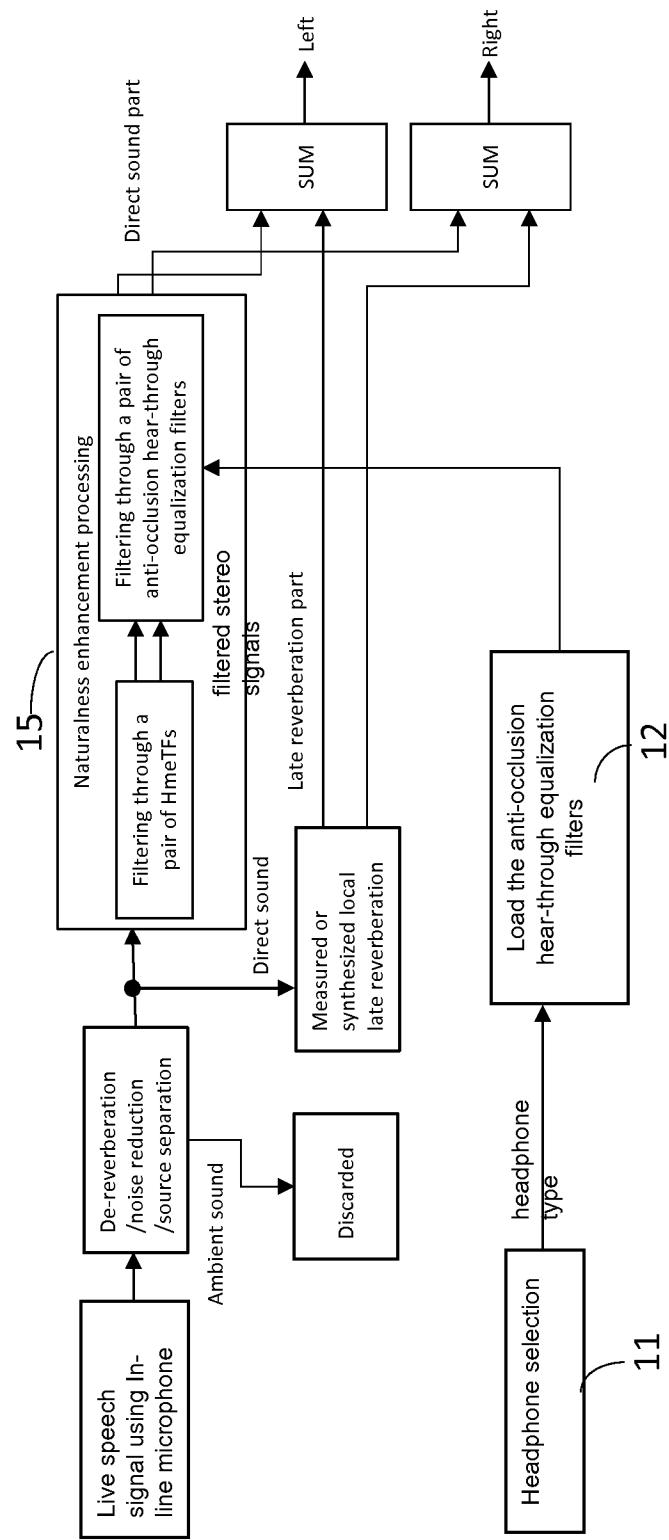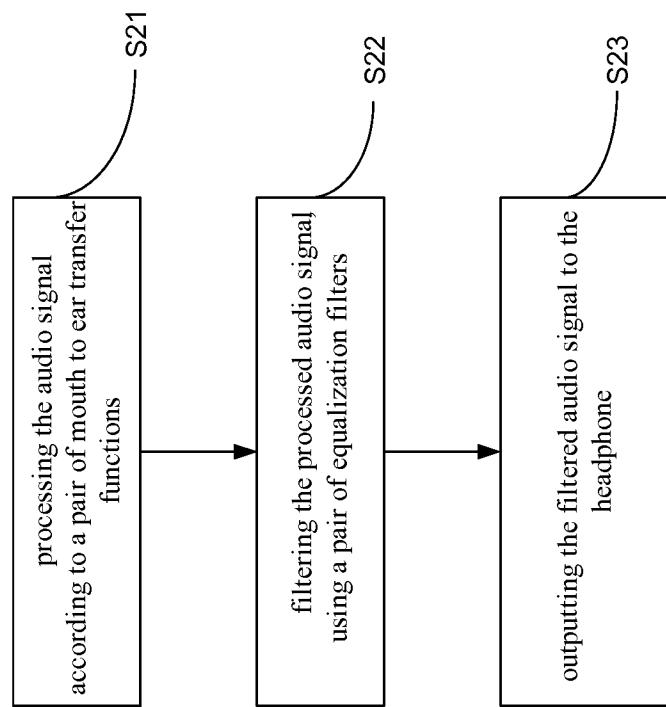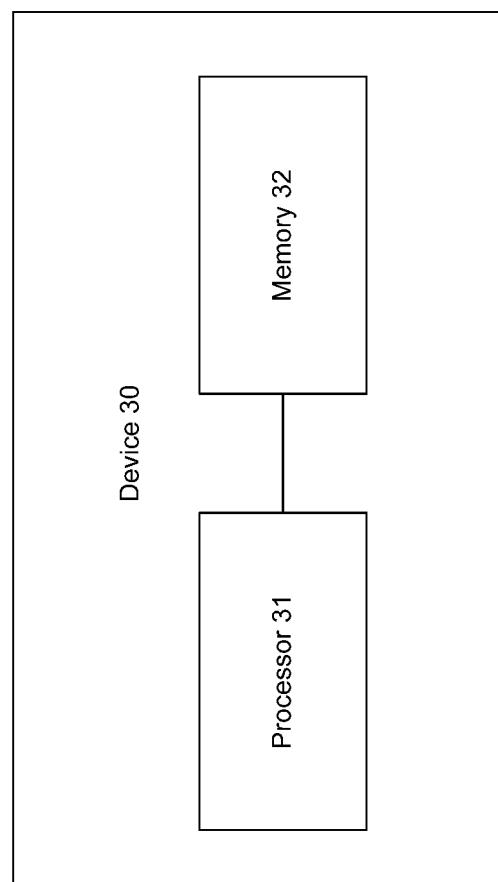
FIG. 14

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04R3/04
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04R

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data, COMPENDEX, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2016/127829 A1 (RING MARTIN DAVID [US] ET AL) 5 May 2016 (2016-05-05) figures 1-3 -----	1,2,4-8, 10-12
X	US 2014/126735 A1 (GAUGER JR DANIEL M [US]) 8 May 2014 (2014-05-08) paragraphs [0038], [0041], [0068] - [0086]; figure 6 -----	1,2,4-8, 10-12

☐ Further documents are listed in the continuation of Box C.   ☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered
  to be of particular relevance
"E" earlier application or patent but published on or after the international
  filing date
"L" document which may throw doubts on priority claim(s) or which is
  cited to establish the publication date of another citation or other
  special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other
  means
"P" document published prior to the international filing date but later than
  the priority date claimed

"T" later document published after the international filing date or priority
  date and not in conflict with the application but cited to understand
  the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be
  considered novel or cannot be considered to involve an inventive
  step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be
  considered to involve an inventive step when the document is
  combined with one or more other such documents, such combination
  being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 October 2019	11/11/2019

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Radomirescu, B-M

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: **3, 9, 13**
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

   see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (April 2005)

FURTHER INFORMATION CONTINUED FROM   PCT/ISA/ 210

Continuation of Box II.2

Claims Nos.: 3, 9, 13

1) In claims 3 and 9, the "measurable frequency range" is neither defined regarding its relevance to the remaining subject-matter of the claims, nor is a parameter unanimously acknowledged in the field to allow the extraction of its inherent features.
2) Regarding the independent claim
13, it is not clear how a computer which processes a program code could be able to output (filtered) audio signals to headphones, since a generic computer does not have all the means necessary to generate audio signals in accordance with the characteristics of the headphones. For example, a generic computer does not include power amplifiers or impedance matching stages to match the output signals to the headphones. Implicitly, when stored on a computer-readable storge medium, a corresponding program code cannot cause the computer to output (filtered) audio signals to the headphones.
According to the Guidelines F-IV 3.9.1, a lack of clarity arises (see also F-IV 3.9.2).
3) The above-mentioned deficiencies are of such a nature that claims 3, 9, 13 cannot be meaningfully searched.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guidelines C-IV, 7.2), should the problems which led to the Article 17(2) declaration be overcome.

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 2016127829	A1	05-05-2016	CN	107005757 A	01-08-2017
			EP	3213527 A1	06-09-2017
			JP	6495448 B2	03-04-2019
			JP	2017533664 A	09-11-2017
			US	2016127829 A1	05-05-2016
			WO	2016069615 A1	06-05-2016
US 2014126735	A1	08-05-2014	CN	105052170 A	11-11-2015
			EP	2915338 A1	09-09-2015
			HK	1216572 A1	18-11-2016
			JP	5956083 B2	20-07-2016
			JP	2015537467 A	24-12-2015
			US	2014126735 A1	08-05-2014
			WO	2014070992 A1	08-05-2014