



(11) **EP 4 571 742 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
18.06.2025 Bulletin 2025/25

(51) International Patent Classification (IPC):
G10L 25/51 ^(2013.01) **G10L 21/0208** ^(2013.01)
G10L 21/0232 ^(2013.01) **G10L 25/18** ^(2013.01)
G10L 25/78 ^(2013.01)

(21) Application number: **23216850.0**

(22) Date of filing: **14.12.2023**

(52) Cooperative Patent Classification (CPC):
G10L 25/51; G10L 21/0208; G10L 21/0232;
G10L 25/18; G10L 25/78

(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 ME MK MT NL
NO PL PT RO RS SE SI SK SM TR**

Designated Extension States:
BA

Designated Validation States:
KH MA MD TN

(72) Inventors:
• **GUJRAL, Rashi Kumra**
5656 AG Eindhoven (NL)
• **PATIL, Ravindra Balasaheb**
5656 AG Eindhoven (NL)
• **GUTTA, Venkata Rama Srinivas**
5656 AG Eindhoven (NL)
• **VALLERI, Vinodh**
5656 AG Eindhoven (NL)

(71) Applicant: **Koninklijke Philips N.V.**
5656 AG Eindhoven (NL)

(74) Representative: **Philips Intellectual Property &
Standards**
High Tech Campus 52
5656 AG Eindhoven (NL)

(54) **PROCESSING AUDIO SIGNALS OF PERSONAL CARE DEVICES**

(57) A mechanism for filtering or attenuating a representation of voice in an audio signal of a personal care device. The audio signal is split or divided into a plurality of audio signal portions. Each audio signal portion is processed to predict whether or not the audio signal portion contains a representation of voice. One or more properties of the audio signal portions identified as containing a representation of voice are used to attenuate a representation of voice in the audio signal.

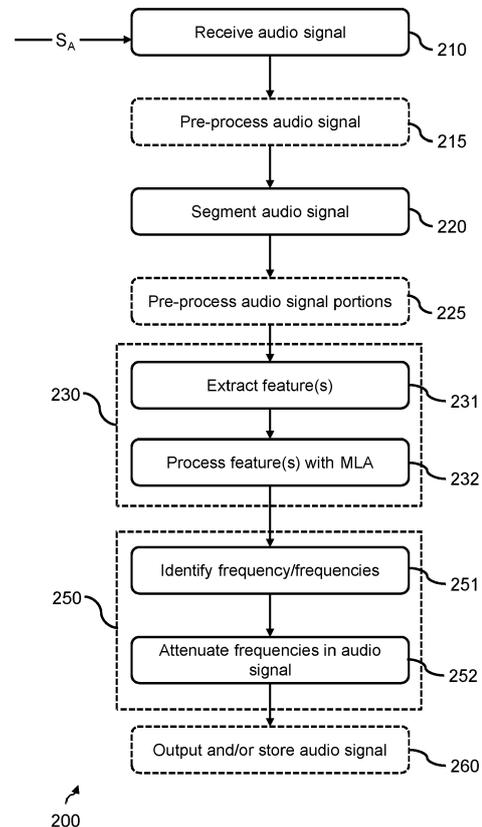


FIG. 2

EP 4 571 742 A1

Description

FIELD OF THE INVENTION

5 **[0001]** The present invention relates to the field of personal care devices, and in particular to the processing of audio signals for personal care devices.

BACKGROUND OF THE INVENTION

10 **[0002]** There is an increasing use of personal care devices globally. Examples of personal care devices are well known to the skilled person, and can include any device usable for improving hygiene of an individual, modifying the appearance of the individual and/or for aiding in the performance of bodily functions of the individual.

[0003] Suitable examples of personal care devices include: toothbrushes, mouthpieces, shavers, hair trimmers, epilators, oral irrigators, intense pulsed light (IPL) treatment devices for hair removal, breast pumps and so on.

15 **[0004]** There is a desire to monitor the performance of a personal care device, e.g., to track the usage of the personal care device, to identify any functional errors in the personal care device or to identify one or more characteristics of the individual on which the personal care device is used.

[0005] One approach for monitoring the performance of the personal care device is to use an audio signal that responds to sound generated by the personal care device. This audio signal may be appropriately processed to track the performance of the personal care device. For instance, certain sounds may be indicative of an error state in the personal care device, a misuse of the personal care device and/or indicative of characteristics of the individual.

[0006] There is therefore a need to improve the fidelity and reliability of any such audio signal of a personal care device.

SUMMARY OF THE INVENTION

25 **[0007]** The invention is defined by the claims.

[0008] According to examples in accordance with an aspect of the invention, there is provided a computer-implemented method for processing an audio signal for a personal care device. The computer-implemented method comprises: receiving the audio signal, wherein the audio signal changes responsive to sound generated by the personal care device and any voices in a vicinity of the personal care device; segmenting the audio signal into a plurality of audio signal portions; processing each audio signal portion using a machine-learning algorithm to identify any voice-containing portions, wherein a voice-containing portion is a portion that is predicted by the machine-learning algorithm to contain a representation of a voice in the vicinity of the personal care device; and if at least one voice-containing portion is identified, performing an attenuation process on the audio signal, to attenuate any representation of the voice in the audio signal, using one or more properties of the at least one identified voice-containing portions.

35 **[0009]** The present disclosure provides a technique for attenuating representations of voice within an audio signal for monitoring noise produced by a personal care device. This can act to both improve a quality of the audio signal, i.e., by attenuating noise resulting from voice(s), with a secondary effect of improved privacy for individuals in the vicinity of the personal care device.

40 **[0010]** The disclosed approach proposes to identify voice-containing portions of an audio signal. Properties or characteristics of the voice-containing portions are then used to attenuate any representation of the voice in the overall audio signal. Thus, the attenuation of the signal is dependent upon the specific use case scenario of the audio signal, rather than employing a global or predefined attenuation technique. This improves the accuracy of attenuating the voice containing portion whilst minimizing a likelihood that sound generated by the personal case device will be attenuated.

45 **[0011]** In some examples, the attenuation process comprises: processing any identified voice-containing portions to identify one or more frequencies of the audio signal attributable to voice; and attenuating frequencies of the audio signal responsive to the identified one or more frequencies.

[0012] This approach acts to attenuate frequencies that are identified as relating or being otherwise associated with a representation of voice in the audio signal. This attenuates voice-containing parts of the audio signal, thereby improving a signal-to-noise ratio of the overall audio signal, with a secondary effect of improved privacy for the individuals providing the voice(s).

50 **[0013]** In some examples, in the attenuation process, attenuating frequencies of the audio signal comprises: defining a frequency filter using the identified one or more frequencies of the audio signal attributable to voice; and applying the frequency filter to the audio signal. This provides a mechanism for directed and specific attenuation of the frequencies attributable to voice present in the audio signal.

55 **[0014]** In some examples, in the attenuation process, processing any identified voice-containing portions comprises: converting the voice-containing portions to the frequency domain using a Fourier-based transform; and identifying, as one of the one or more frequencies, a threshold frequency by processing the voice containing portions in the frequency domain.

[0015] This provides a technique for accurate identification of one or more frequencies attributable to voice. In particular, by converting the audio signal portion into the frequency domain using a Fourier-based transform, frequencies present in voice-containing portions can be readily identified - as it can be assumed that such frequencies are attributable to a vocal representation within the audio signal portion.

5 **[0016]** The threshold frequency may represent a cut-off frequency for a low-pass filter, e.g., a Butterworth or Chebyshev filter. The attenuation process may comprise applying a low-pass filter with the identified threshold frequency as the cut-off frequency for the low-pass filter.

[0017] The attenuation process may further comprise amplifying frequencies of the audio signal that are not attributable to voice. This improves the signal-to-noise ratio of the audio signal, by amplifying those portions that are predicted to not result from undesirable voices produced by an individual.

10 **[0018]** The machine-learning algorithm may be or comprise a balanced random forest classifier. It has been herein identified that this form of classifier is particularly suited and accurate to performing a task of identifying whether or not an audio signal portion contains a representation of voice (i.e., is a voice-containing audio signal portion).

[0019] In some examples, the step of processing each audio signal portion comprises, for each audio signal portion: processing the audio signal portion using a feature extraction process to extract one or features of the audio signal portion; and inputting the extracted one or more features into the machine-learning algorithm to determine whether or not the audio signal portion is a voice-containing portion.

[0020] It has been identified that certain features of an audio signal portion improve the accuracy of predicting whether or not the audio signal portion is a voice-containing audio signal portion.

20 **[0021]** The feature extraction process may comprise a Mel Frequency Cepstral Coefficient (MFCC) process, such that the one or more features comprises one or more Mel Frequency Cepstral Coefficients of the audio signal portion.

[0022] The feature extraction process may comprise a zero-crossing rate process, such that the one or more features comprises a zero-crossing rate of the audio signal portion.

[0023] In some examples, the one or more features comprises one or more of: a short time energy, a spectral flatness measurements and/or an auto correlation value.

25 **[0024]** The plurality of audio signal portions may temporally abut one another. Thus, there may be no overlap between the audio signal portions, but rather they may be temporally positioned end-to-end.

[0025] In preferred examples, the length of each audio signal portion is no greater than 20 ms. For instance, the length of each audio signal portion may be no greater than 15 ms, e.g., 10 ms.

30 **[0026]** In some examples, the personal care device is a laser hair removal device. A laser hair removal device will produce a tone or noise at frequent/fixed intervals, such that it is possible to more accurately distinguish portions containing noise produced by the personal care device from voice-containing portions.

[0027] There is also provided a computer program product comprising computer program code means which, when executed on a computing device having a processing system, cause the processing system to perform all of the steps of any herein disclosed method.

35 **[0028]** There is also provided a processing system for processing an audio signal for a personal care device, the processing system being configured to: receive the audio signal, wherein the audio signal changes responsive to sound generated by the personal care device and any voices in a vicinity of the personal care device; segment the audio signal into a plurality of audio signal portions; process each audio signal portion using a machine-learning algorithm to identify any voice-containing portions, wherein a voice-containing portion is a portion that is predicted by the machine-learning algorithm to contain a representation of a voice in the vicinity of the personal care device; and if at least one voice-containing portion is identified, perform an attenuation process on the audio signal, to attenuate any representation of the voice in the audio signal, using the at least one identified voice-containing portions.

45 **[0029]** These and other aspects of the invention will be apparent from and elucidated with reference to the embodiment(s) described hereinafter.

BRIEF DESCRIPTION OF THE DRAWINGS

50 **[0030]** For a better understanding of the invention, and to show more clearly how it may be carried into effect, reference will now be made, by way of example only, to the accompanying drawings, in which:

- Fig. 1 illustrates a workflow in which embodiments may be employed;
- Fig. 2 is a flowchart illustrating a proposed method;
- Fig. 3 is a waveform illustrating an audio signal before being processed by a proposed method; and
- 55 Fig. 4 is a waveform illustrating the audio signal after being processed by the proposed method.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0031] The invention will be described with reference to the Figures.

[0032] It should be understood that the detailed description and specific examples, while indicating exemplary embodiments of the apparatus, systems and methods, are intended for purposes of illustration only and are not intended to limit the scope of the invention. These and other features, aspects, and advantages of the apparatus, systems and methods of the present invention will become better understood from the following description, appended claims, and accompanying drawings. It should be understood that the Figures are merely schematic and are not drawn to scale. It should also be understood that the same reference numerals are used throughout the Figures to indicate the same or similar parts.

[0033] The invention provides a mechanism for filtering or attenuating a representation of voice in an audio signal of a personal care device. The audio signal is split or divided into a plurality of audio signal portions. Each audio signal portion is processed to predict whether or not the audio signal portion contains a representation of voice. One or more properties of the audio signal portions identified as containing a representation of voice are used to attenuate a representation of voice in the audio signal.

[0034] Fig. 1 illustrates a workflow 10 in which proposed embodiments may be employed.

[0035] An audio signal S_A is detected at or in the vicinity of a personal care device 100, e.g., (as illustrated) a laser hair removal device. The audio signal may, for instance, be generated by a microphone 110, a microphone array or other noise-sensitive electronic device that is positioned or located to detect sound generated by the personal care device. The audio signal S_A captures sound 105 or noise produced by the personal care device 100 during its use. In this way, the audio signal S_A will contain a representation of any sound 105 or noise produced by the personal care device during operational use.

[0036] The audio signal S_A may be processed by a processing system 150, e.g., to monitor the performance of the personal care device. For instance, the processing system 150 may wish to track the usage of the personal care device, to identify any functional errors in the personal care device or to identify one or more characteristics of the individual on which the personal care device is used.

[0037] The audio signal S_A does not need to be processed during the use of the personal care device. For instance, the audio signal S_A may be stored and/or collated with other audio signals before being processed at a later point in time.

[0038] The audio signal S_A or parts thereof may be stored by the processing system 150, e.g., as a result of the processing or in advance of the processing. For instance, the audio signal S_A may be processed to identify any portions of the audio signal that identify or indicate a predicted error in usage or operation of the personal care device for later analysis and/or assessment.

[0039] Although illustrated as separate elements, in practice, the microphone 110 and/or processing system 150 may form part of the personal care device, e.g., be integrated as an aspect of the personal care device. In some (alternative) examples, the microphone 110 is formed as a separate element, e.g., as a microphone of a smartphone or other electronic device in the proximity (e.g., within 2m or within 1m) of the personal health care device.

[0040] The present disclosure recognizes that the audio signal S_A will be responsive to any sound(s) in the vicinity of the personal care device, i.e., as the microphone will be indiscriminatory.

[0041] In particular, the audio signal S_A will be responsive to any voices 195 in the vicinity of the personal care device (e.g., produced by one or more individuals 190 talking whilst the personal care device is being used). The presence of these voices in the audio signal can significantly affect an accuracy of the audio signal with respect to the sound generated by the personal care device, and increase a difficulty of monitoring the performance of the personal care device. Moreover, there is a further consideration of privacy for the individual(s) producing the voice, e.g., should the audio signal or portions thereof be stored for later analysis.

[0042] The present disclosure provides a mechanism for attenuating the representation of the/any voice in the audio signal, to mitigate the impact of these disadvantages.

[0043] Fig. 2 is a flowchart illustrating a computer-implemented method 200 for processing an audio signal for a personal care device. The computer-implemented method may be performed by the processing system previously mentioned.

[0044] The method 200 comprises a step 210 of receiving the audio signal S_A . The audio signal changes responsive to sound generated by the personal care device and any voices in a vicinity of the personal care device.

[0045] In some examples, step 210 comprises generating the audio signal using a microphone or other noise sensitive arrangement. In other examples, step 210 comprises receiving the audio signal from the microphone or other noise sensitive arrangement. In yet other examples, step 210 comprises retrieving or receiving the audio signal from a memory or storage device.

[0046] The method also comprises a step 220 of segmenting the audio signal into a plurality of audio signal portions. The audio signal portions may be configured to abut one another, e.g., to not overlap, although this is not essential. The length of each portion is less than the total length of the audio signal. The length of each audio signal portion may, for instance, be no greater than 50 ms, e.g., no greater than 20 ms, e.g., 10 ms.

[0047] The method also comprises a step 230 of processing each audio signal portion using a machine-learning

algorithm to identify any voice-containing portions. A voice-containing portion is a portion that is predicted by the machine-learning algorithm to contain a representation of a voice in the vicinity of the personal care device.

[0048] This step recognizes that a voice of an individual (or voices or individuals) is/are unlikely to fill the entirety of the audio signal. Rather, different sections or segments of the audio signal will contain a representation of voice.

[0049] The method 200 also comprises, if at least one voice-containing portion is identified, performing an attenuation process 250 on the audio signal, to attenuate any representation of the voice in the audio signal, using one or more properties of the at least one identified voice-containing portions.

[0050] Thus, the method 200 may comprise a step 240 of determining whether or not any voice-containing portions have been identified. Responsive to a positive determination, the method performs the attenuation process. Otherwise, the method may restart or end.

[0051] The attenuation process 250 may comprise a step 251 of processing any identified voice-containing portions to identify one or more frequencies of the audio signal attributable to voice and a step 252 of attenuating frequencies of the audio signal responsive to the identified one or more frequencies.

[0052] In this way, the voice-containing portions are processed to identify and attenuate frequencies attributed to voice within the audio signal. This facilitates bespoke identification of vocal frequencies within the audio signal, thereby reducing a risk of other frequencies being attenuated. These other frequencies may carry important information about the operation of the personal care device.

[0053] Step 252 may be performed by defining a frequency filter using the identified one or more frequencies of the audio signal attributable to voice; and applying the frequency filter to the audio signal.

[0054] For instance, the frequency filter may comprise a plurality of band-stop filters (e.g., notch filters), each band-stop filter being centered at a respective one of the identified one or more frequencies of the audio signal attributable to voice. This advantageously tunes the filtering to the specific vocal frequencies identifiable in the audio signal, to reduce attenuation of non-vocal elements of the audio signal.

[0055] As another example, frequency filter may comprise one or more band-stop filters (e.g., notch filters), each band-stop filter being centered to attenuate a respective set or cluster of one or more of the identified one or more frequencies of the audio signal attributable to voice. This advantageously tunes the filtering to the specific vocal frequencies identifiable in the audio signal, to reduce attenuation of non-vocal elements of the audio signal.

[0056] As another example, the frequency filter may be a low-pass filter, in which the cut-off frequency of the low-pass filter is defined by the lowest of the identified one or more frequencies of the audio signal attributable to voice.

[0057] Other suitable examples will be apparent to the skilled person, upon being taught to attenuate frequencies attributable to voice from an audio signal based on one or more identified frequencies in voice-containing portions of the audio signal.

[0058] As one example, step 251 may comprise converting the voice-containing portions to the frequency domain using a Fourier-based transform (such as an FFT); and identifying, as one of the one or more frequencies, a threshold frequency by processing the voice containing portions in the frequency domain.

[0059] The threshold frequency may, for instance, represent the lowest frequency that breaches some predetermined amplitude threshold. This threshold frequency may, for instance, be used to define a cut-off frequency for a low-pass filter or a center frequency of a band-stop filter.

[0060] In some examples, a plurality of threshold frequencies is identified, each threshold frequency representing an identified frequency that breaches some predetermined amplitude threshold. A band-stop filter may then be centered at a respective one of the identified threshold frequencies - or designed to cover one or more ranges of the identified threshold frequencies.

[0061] The attenuation process 250 may further comprise a step 253 of amplifying frequencies of the audio signal that are not attributable to voice. This improves the signal-to-noise ratio of the audio signal.

[0062] As a full working example, the attenuation process may comprise performing a Fourier-based transform (e.g., a fast Fourier transform FFT) over the voice-containing portions of the audio signal to produce a frequency spectrum of the voice-containing portions. A threshold frequency may then be calculated using the frequency spectrum, e.g., further based upon a desired sensitivity of the algorithm. For instance, the threshold frequency may represent the lowest frequency that breaches some predetermined amplitude threshold. The attenuation process may then perform the same Fourier-based transform (e.g., FFT) on the entire audio signal to produce a frequency spectrum of the audio signal. A mask may then be determined by comparing the frequency spectrum of the audio signal to the threshold frequency. The mask represents the (desired) suppression frequency and time range to be smoothed and is computed based on a predetermined given threshold. Optionally, the mask is smoothed with a predetermined filter (e.g., over frequency and time), such as one that employs a spectral subtraction technique. The (optionally smoothed) mask is applied to the frequency spectrum of the audio signal to effectively attenuate frequencies attributable to voice. The reverse of the Fourier-based transform is performed on the resulting frequency spectrum to produce the processed audio signal that is to be output by the attenuation process.

[0063] Fig. 3 illustrates an example of an audio signal 300 that changes responsive to sound generated by the personal

care device and any voices in a vicinity of the personal care device. In this example, the personal care device is a laser hair removal device.

[0064] Fig. 4 illustrates the same example of the audio signal 300 after it has been processed by performing the previously described method 200, described with reference to Fig. 2. The frequencies corresponding to any voice have been attenuated, and other frequencies have been amplified to improve the signal-to-noise ratio of the audio signal.

[0065] In the above-described examples, the attenuation process identifies one or more frequencies in voice-containing audio signal portions, and attenuates those frequencies. However, alternative properties of the voice containing audio signal portions can be used in the attenuation process (e.g., in addition to or instead of the frequencies).

[0066] As one example, the attenuation process may identify an average amplitude of the voice-containing audio signal portions, and may attenuate (in the audio signal) any amplitude below this average amplitude. This improves the signal-to-noise ratio, especially if it is known or defined that the representation of noise produced by the personal care device will have a higher amplitude than any representation of voice in the audio signal.

[0067] It has previously been explained how each audio signal portion is processed by a machine-learning algorithm to predict, for each audio signal portion, whether that audio signal portion is a voice-containing portion.

[0068] A machine-learning algorithm is any self-training algorithm that processes input data in order to produce or predict output data. Here, the input data comprises an audio signal portion (or, preferably, features derived therefrom) and the output data comprises a prediction or classification of whether the audio signal portion contains a voice.

[0069] Suitable machine-learning algorithms for being employed in the present invention will be apparent to the skilled person. Examples of suitable machine-learning algorithms include decision tree algorithms and artificial neural networks. Other machine-learning algorithms such as logistic regression, support vector machines or Naive Bayesian models are suitable alternatives.

[0070] The structure of an artificial neural network (or, simply, neural network) is inspired by the human brain. Neural networks are comprised of layers, each layer comprising a plurality of neurons. Each neuron comprises a mathematical operation. In particular, each neuron may comprise a different weighted combination of a single type of transformation (e.g. the same type of transformation, sigmoid etc. but with different weightings). In the process of processing input data, the mathematical operation of each neuron is performed on the input data to produce a numerical output, and the outputs of each layer in the neural network are fed into the next layer sequentially. The final layer provides the output.

[0071] A decision tree algorithm processes input data through a tree of nodes. In the tree of nodes, each successive node splits into two or more further nodes until reaching a terminal or end node. When performing the decision tree algorithm using the tree of nodes, at each node, a decision is made as to which further node to move to next based on the input data. The end node defines the outcome of the decision tree algorithm, and therefore the machine-learning method.

[0072] Methods of training a machine-learning algorithm are well known. Typically, such methods comprise obtaining a training dataset, comprising training input data entries and corresponding training output data entries.

[0073] For some machine-learning algorithms training is performed by applying initialized machine-learning algorithm to each input data entry to generate predicted output data entries. An error between the predicted output data entries and corresponding training output data entries is used to modify the machine-learning algorithm. This process can be repeated until the error converges, and the predicted output data entries are sufficiently similar (e.g. $\pm 1\%$) to the training output data entries. This is commonly known as a supervised learning technique.

[0074] For example, where the machine-learning algorithm is formed from a neural network, (weightings of) the mathematical operation of each neuron may be modified until the error converges. Known methods of modifying a neural network include gradient descent, backpropagation algorithms and so on.

[0075] Other approaches for training machine-learning algorithms (e.g., decision trees) are known in the art. For instance, decision trees are often trained using a decision tree builder or learning techniques, such as those set out by Suthaharan, Shan, and Shan Suthaharan. "Decision tree learning." *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning* (2016): 237-269 or Ruggieri, Salvatore. "Yadt: Yet another decision tree builder." 16th IEEE International Conference on Tools with Artificial Intelligence. IEEE, 2004.

[0076] The training input data entries correspond to example audio signal portions (or, preferably, features derived therefrom). The training output data entries correspond to a classification of whether or not the audio signal portion contains a representation of voice.

[0077] The training output data entries may be readily generated by an appropriately skilled or trained individual labelled portions of an example audio signal that contain a representation of voice. The example audio signal may be divided into audio signal portions, and an appropriate label assigned to each portion dependent upon whether or not the audio signal portion falls within one of the labelled portions of the example audio signal (e.g., by comparing timestamps of the audio signal portion to the labelled portion(s)).

[0078] For the purposes of the present disclosure, it has been identified that a balanced random forest classifier is particularly advantageous/suited for predicting whether or not an audio signal portion is a voice-containing portion.

[0079] More particularly, the balanced random forest classifier has been experimentally identified as having superior performance for the present application when compared to other forms of binary classification algorithm such as Support

Vector Machine (SVM), logistic regression, and Extreme Gradient Boosting (XGBoost). The testing accuracy obtained for SVM, logistic regression, XGBoost were 68%, 57%, and 70% respectively. The best accuracy was given by Balanced Random Forest Classifier.

[0080] The use of a balanced random forest classifier also overcomes or avoids any problem of imbalance distribution of positive and negative classes.

[0081] To improve the accuracy of predicting whether or not an audio signal portion is a voice-containing portion - it may be preferable to extract one or more acoustic features from the audio signal portion before processing using the machine-learning algorithm.

[0082] Accordingly (and referring back to Fig. 2), in preferred examples, the step 230 of processing each audio signal portion comprises performing, for each audio signal portion, a sub-step 231 and a sub-step 232. Sub-step 231 comprises processing the audio signal portion using a feature extraction process to extract one or features of the audio signal portion. Sub-step 232 comprises the extracted one or more features into the machine-learning algorithm to determine whether or not the audio signal portion is a voice-containing portion.

[0083] The skilled person would readily appreciate how to appropriately train a machine-learning algorithm to process the extracted features of an audio signal portion to predict whether or not the audio signal portion is a voice-containing portion. Such techniques comprise may comprise defining or extracting similar features from labelled examples of audio signal portions.

[0084] A number of examples of suitable features for extraction from an audio signal portion are hereafter described. Embodiments may employ or extract any one or more of the hereafter described features.

[0085] In some examples, the feature extraction process comprises a Mel Frequency Cepstral Coefficient process, such that the one or more features comprises one or more Mel Frequency Cepstral Coefficients of the audio signal portion. Approaches for deriving or extracting Mel Frequency Cepstral Coefficients (MFCCs) are well-established in the art, for instance, as set out in Abdul, Zrar Kh, and Abdulbasit K. Al-Talabani. "Mel Frequency Cepstral Coefficient and its applications: A Review." IEEE Access (2022). The skilled person would be readily capable of applying such approaches to (the data of) an audio signal portion.

[0086] In some examples, only a restricted number of the MFCCs are used or identified, as there is a diminishing return in increasing the number of MFCCs used in identifying whether or not the audio signal portion contains a representation of voice. As a working example, the first 13 MFCCs may be included in the one or more features, as these coefficients contain the majority or most of the information for performing speech analysis.

[0087] In some examples, the feature extraction process comprises a zero-crossing rate process, such that the one or more features comprises a zero-crossing rate of the audio signal portion. The zero crossing rate may effectively define how many times the amplitude of the signal (in the audio signal portion) crosses zero within the length of the audio signal portion. The following equation defines one mechanism for determining a zero-rate crossing (Z_j) of a signal $x(i)$.

$$Z_j = \sum_{i=N(j-1)+1}^{jN} |\text{sgn}[x(i)] - \text{sgn}[x(i-1)]| \quad (1)$$

[0088] The function $\text{sgn}[\cdot]$ is the sign function that identifies the sign of a value, where the function $\text{sgn}[\cdot]$ outputs a first predetermined value (e.g., 1) for a value greater than 0 and a second predetermined value (being 0 minus the first predetermined value, e.g., -1) for a value less than 0.

[0089] It is noted that, in general, the zero-crossing rate of an audio signal portion will be lower for voice-containing audio signal portions that do not contain a representation of voice.

[0090] In some examples, the feature extraction process is configured to identify, as one of the one or more features, a short time energy of the audio signal portion. Thus, the one or more features may comprise a short time energy of the audio signal portion. It is noted that, in general, the value of the short-time energy of the audio signal portion will be higher for voice-containing audio signal portions that do not contain a representation of voice.

[0091] One approach for determining the short time energy of an audio signal portion is to make use of the following equation:

$$E_n = \sum_{m=0}^{N-1} x_n^2(m) \quad (2)$$

where N is the number of samples for an audio signal portion, $x_n(m)$ is the m -th sample of the audio signal portion and E_n is the short time energy of the audio signal portion.

[0092] In some examples, the feature extraction process is configured to identify, as one of the one or more features, a

Spectral Flatness Measurement (SFM). Thus, the one or more features may comprise a spectral flatness measurement of the audio signal portion. It is noted that, in general, the value of the spectral flatness measurement of the audio signal portion will be lower for voice-containing audio signal portions that do not contain a representation of voice.

[0093] One approach for determining a spectral flatness measurement SFM(m) of an audio signal portion m is to make use of the following equation:

$$SFM(m) = \frac{GM(PS(M))}{AM(PS(M))} \quad (3)$$

where PS(M) is the power spectrum of the audio signal portion m, e.g., calculated or determined using a Fourier-based transform such as an FFT. GM(PS(M)) is the geometric mean of the power spectrum PS(M) and AM(PS(M)) is the arithmetic mean of the power spectrum PS(M).

[0094] In some examples, the feature extraction process is configured to perform an auto correlation function on an audio signal portion to identify an auto correlation value C1 of the audio signal portion. Thus, the one or more features may comprise an auto correlation value C1 of the audio signal portion. It is noted that, in general, the value of the autocorrelation value C1 will approach unity (i.e., 1) for voice-containing portions and will approach 0 for portions that do not contain a representation of voice.

[0095] One approach for determining an auto correlation value C1 is to perform the following function:

$$C1 = \frac{\sum_{n=1}^N s(n)s(n-1)}{\sqrt{(\sum_{n=1}^N s^2(n))(\sum_{n=0}^N s^2(n))}} \quad (4)$$

where s(n) is the n-th (sequential) sample of the audio signal portion, where there are N sequential samples in the audio signal portion.

[0096] Additional filtering and/or signal processing or pre-processing may be performed on the audio signal, e.g., prior to performing step 220, and/or the audio signal portions (e.g., prior to processing using the machine-learning algorithm).

[0097] As an example, the method 200 may further comprise a step 215 of performing signal pre-processing on the audio signal. Step 215 may, for instance, comprise processing the audio signal using one or more wavelet filters, such as one or more orthogonal wavelet filters.

[0098] As another example, the method 200 may further comprise a step 225 of performing signal pre-processing on each audio signal portion. Step 225 may, for instance, comprise processing each audio signal portion using one or more wavelet filters, examples of which are known in the art, such as orthogonal wavelet filters.

[0099] As previously mentioned, the processed audio signal may be output for further processing and/or undergo further processing as part of the method 200. In some examples, the processed audio signal may be stored (e.g., in a storage or memory device) and/or transmitted for further processing (e.g., to a further processing device).

[0100] Thus, the method 200 may further comprise a step 260 of storing and/or outputting the audio signal.

[0101] The skilled person would be readily capable of developing a processing system for carrying out any herein described method. Thus, each step of the flow chart may represent a different action performed by a processing system, and may be performed by a respective module of the processing system.

[0102] Embodiments may therefore make use of a processing system. The processing system can be implemented in numerous ways, with software and/or hardware, to perform the various functions required. A processor is one example of a processing system which employs one or more microprocessors that may be programmed using software (e.g., microcode) to perform the required functions. A processing system may however be implemented with or without employing a processor, and also may be implemented as a combination of dedicated hardware to perform some functions and a processor (e.g., one or more programmed microprocessors and associated circuitry) to perform other functions.

[0103] Examples of processing system components that may be employed in various embodiments of the present disclosure include, but are not limited to, conventional microprocessors, application specific integrated circuits (ASICs), and field-programmable gate arrays (FPGAs).

[0104] In various implementations, a processor or processing system may be associated with one or more storage media such as volatile and non-volatile computer memory such as RAM, PROM, EPROM, and EEPROM. The storage media may be encoded with one or more programs that, when executed on one or more processors and/or processing systems, perform the required functions. Various storage media may be fixed within a processor or processing system or may be transportable, such that the one or more programs stored thereon can be loaded into a processor or processing system.

[0105] It will be understood that disclosed methods are preferably computer-implemented methods. As such, there is also proposed the concept of a computer program comprising code means for implementing any described method when said program is run on a processing system, such as a computer. Thus, different portions, lines or blocks of code of a computer program according to an embodiment may be executed by a processing system or computer to perform any herein described method.

[0106] There is also proposed an audio processing system comprising the processing system and a microphone configured to generate the audio signal.

[0107] There is also proposed a personal care device system comprising the audio processing system and a personal care device, such as a laser hair removal device.

[0108] There is also proposed a non-transitory storage medium that stores or carries a computer program or computer code that, when executed by a processing system, causes the processing system to carry out any herein disclosed method.

[0109] In some alternative implementations, the functions noted in the block diagram(s) or flow chart(s) may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved.

[0110] Variations to the disclosed embodiments can be understood and effected by those skilled in the art in practicing the claimed invention, from a study of the drawings, the disclosure and the appended claims. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage.

[0111] In the claims, the word "comprising" does not exclude other elements or steps, and the indefinite article "a" or "an" does not exclude a plurality. If the term "adapted to" is used in the claims or description, it is noted the term "adapted to" is intended to be equivalent to the term "configured to". If the term "arrangement" is used in the claims or description, it is noted the term "arrangement" is intended to be equivalent to the term "system", and vice versa.

[0112] A single processor or other unit may fulfill the functions of several items recited in the claims. If a computer program is discussed above, it may be stored/distributed on a suitable medium, such as an optical storage medium or a solid-state medium supplied together with or as part of other hardware, but may also be distributed in other forms, such as via the Internet or other wired or wireless telecommunication systems.

Claims

1. A computer-implemented method for processing an audio signal for a personal care device, the computer-implemented method comprising:

receiving the audio signal, wherein the audio signal changes responsive to sound generated by the personal care device and any voices in a vicinity of the personal care device;
 segmenting the audio signal into a plurality of audio signal portions;
 processing each audio signal portion using a machine-learning algorithm to identify any voice-containing portions, wherein a voice-containing portion is a portion that is predicted by the machine-learning algorithm to contain a representation of a voice in the vicinity of the personal care device; and
 if at least one voice-containing portion is identified, performing an attenuation process on the audio signal, to attenuate any representation of the voice in the audio signal, using one or more properties of the at least one identified voice-containing portions.

2. The computer-implemented method of claim 1, wherein the attenuation process comprises:

processing any identified voice-containing portions to identify one or more frequencies of the audio signal attributable to voice; and
 attenuating frequencies of the audio signal responsive to the identified one or more frequencies.

3. The computer-implemented method of claim 2, wherein, in the attenuation process, attenuating frequencies of the audio signal comprises:

defining a frequency filter using the identified one or more frequencies of the audio signal attributable to voice; and
 applying the frequency filter to the audio signal.

4. The computer-implemented method of claim 2 or 3, wherein, in the attenuation process, processing any identified voice-containing portions comprises:

converting the voice-containing portions to the frequency domain using a Fourier-based transform; and identifying, as one of the one or more frequencies, a threshold frequency by processing the voice containing portions in the frequency domain.

- 5 **5.** The computer-implemented method of any of claims 1 to 4, wherein the attenuation process further comprises amplifying frequencies of the audio signal that are not attributable to voice.
6. The computer-implemented method of any of claims 1 to 5, wherein the machine-learning algorithm comprises a balanced random forest classifier.
- 10 **7.** The computer-implemented method of any of claims 1 to 6, wherein the step of processing each audio signal portion comprises, for each audio signal portion:
- 15 processing the audio signal portion using a feature extraction process to extract one or features of the audio signal portion; and
inputting the extracted one or more features into the machine-learning algorithm to determine whether or not the audio signal portion is a voice-containing portion.
- 20 **8.** The computer-implemented method of claim 7, wherein the feature extraction process comprises a Mel Frequency Cepstral Coefficient process, such that the one or more features comprises one or more Mel Frequency Cepstral Coefficients of the audio signal portion.
- 25 **9.** The computer-implemented method of claim 7 or 8, wherein the feature extraction process comprises a zero-crossing rate process, such that the one or more features comprises a zero-crossing rate of the audio signal portion.
- 30 **10.** The computer-implemented method of any of claims 7 to 9, wherein the one or more features comprises one or more of: a short time energy, a spectral flatness measurements and/or an auto correlation value.
- 35 **11.** The computer-implemented method of any of claims 1 to 10, wherein the plurality of audio signal portions temporally about one another.
- 40 **12.** The computer-implemented method of any of claims 1 to 11, wherein the length of each audio signal portion is no greater than 20 ms.
- 45 **13.** The computer-implemented method of any of claims 1 to 12, wherein the personal care device is a laser hair removal device.
- 50 **14.** A computer program product comprising computer program code means which, when executed on a computing device having a processing system, cause the processing system to perform all of the steps of the method according to any of claims 1 to 13.
- 55 **15.** A processing system for processing an audio signal for a personal care device, the processing system being configured to:
- receive the audio signal, wherein the audio signal changes responsive to sound generated by the personal care device and any voices in a vicinity of the personal care device;
segment the audio signal into a plurality of audio signal portions;
process each audio signal portion using a machine-learning algorithm to identify any voice-containing portions, wherein a voice-containing portion is a portion that is predicted by the machine-learning algorithm to contain a representation of a voice in the vicinity of the personal care device; and
if at least one voice-containing portion is identified, perform an attenuation process on the audio signal, to attenuate any representation of the voice in the audio signal, using the at least one identified voice-containing portions.

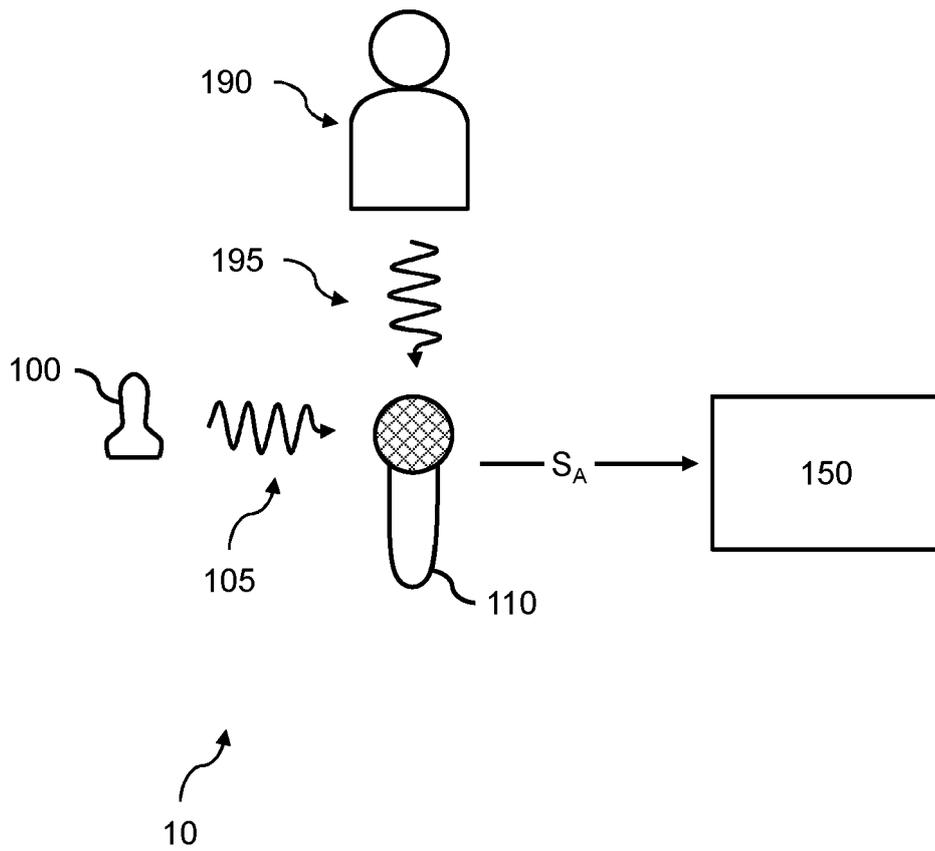


FIG. 1

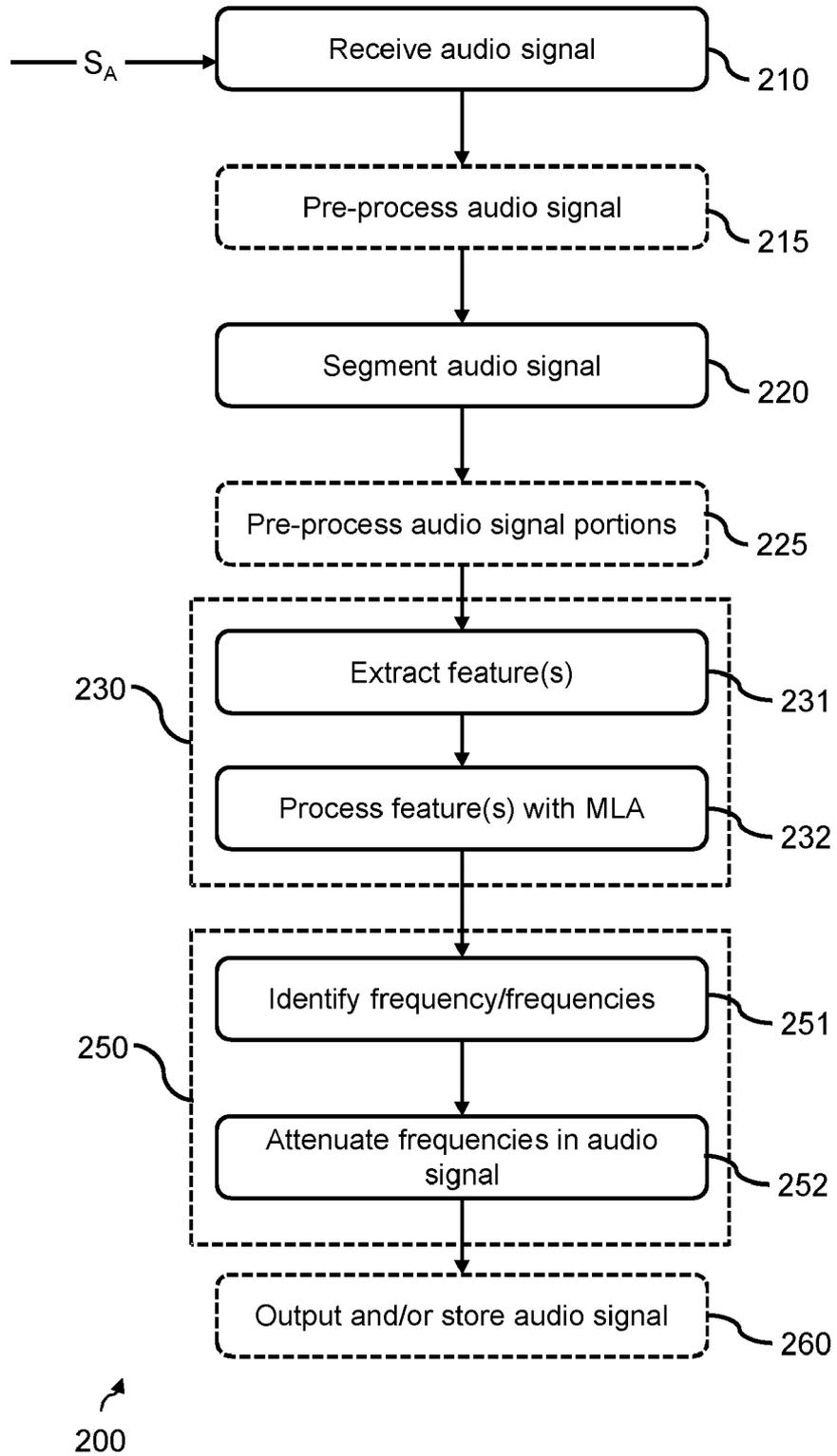


FIG. 2

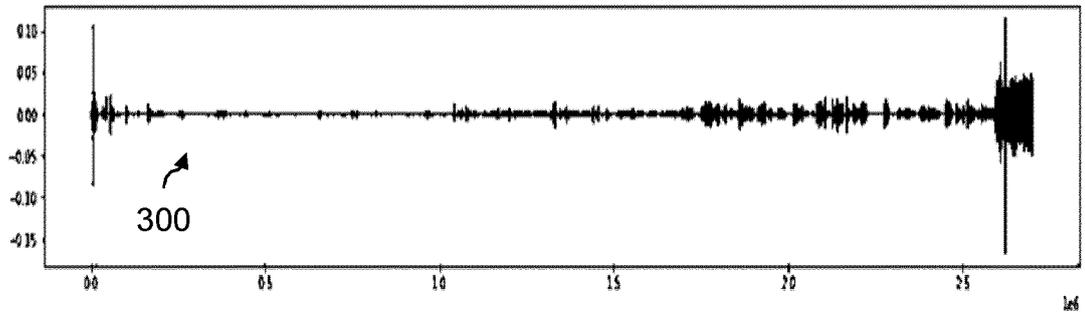


FIG. 3

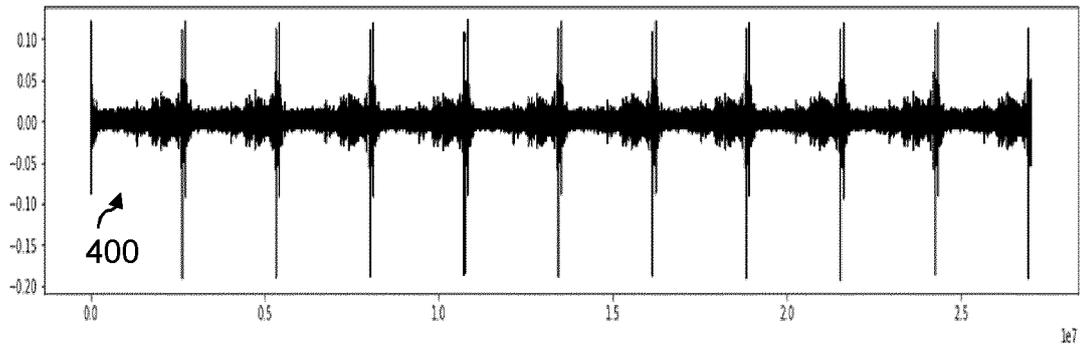


FIG. 4



EUROPEAN SEARCH REPORT

Application Number

EP 23 21 6850

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2023/317086 A1 (VIRTANEN TUOMAS [FI] ET AL) 5 October 2023 (2023-10-05) * paragraph [0024] * * paragraph [0033] * * paragraph [0036] - paragraph [0037] * * paragraph [0039] * * paragraph [0002] * -----	1-15	INV. G10L25/51 G10L21/0208 ADD. G10L21/0232 G10L25/18 G10L25/78
A	US 2022/358954 A1 (IRAVANTCHI YASHA [US] ET AL) 10 November 2022 (2022-11-10) * paragraph [0036] * -----	6	
A	EP 4 226 883 A1 (KONINKLIJKE PHILIPS NV [NL]) 16 August 2023 (2023-08-16) * paragraph [0020] * * paragraph [0080] - paragraph [0081] * -----	9,10	
			TECHNICAL FIELDS SEARCHED (IPC)
			G10L
The present search report has been drawn up for all claims			
Place of search The Hague		Date of completion of the search 25 April 2024	Examiner Stan, Guy-Bart
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	

4
EPO FORM 1503 03.82 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 23 21 6850

5

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.

25-04-2024

10

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2023317086 A1	05-10-2023	CA 3194165 A1	17-03-2022
		EP 4211687 A1	19-07-2023
		US 2023317086 A1	05-10-2023
		WO 2022053742 A1	17-03-2022

US 2022358954 A1	10-11-2022	US 2022358954 A1	10-11-2022
		WO 2022235748 A1	10-11-2022

EP 4226883 A1	16-08-2023	EP 4226883 A1	16-08-2023
		WO 2023156174 A1	24-08-2023

15

20

25

30

35

40

45

50

55

EPO FORM P0459

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