

(51) International Patent Classification:  
G10L 21/02 (2006.01)(21) International Application Number:  
PCT/US2012/024372(22) International Filing Date:  
8 February 2012 (08.02.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
61/441,611 10 February 2011 (10.02.2011) US(71) Applicant (for all designated States except US): **DOLBY LABORATORIES LICENSING CORPORATION** [US/US]; 100 Potrero Avenue, San Francisco, California 94103-4813 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **DICKINS, Glenn N.** [AU/AU]; Level 3, 35 Mitchell Street, McMahon's Point, Sydney, New South Wales NSW 2060 (AU).(74) Agents: **DOLBY LABORATORIES, INC.** et al.; Intellectual Property Group, 999 Brannan Street, San Francisco, California 94103-4938 (US).

(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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR,

KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

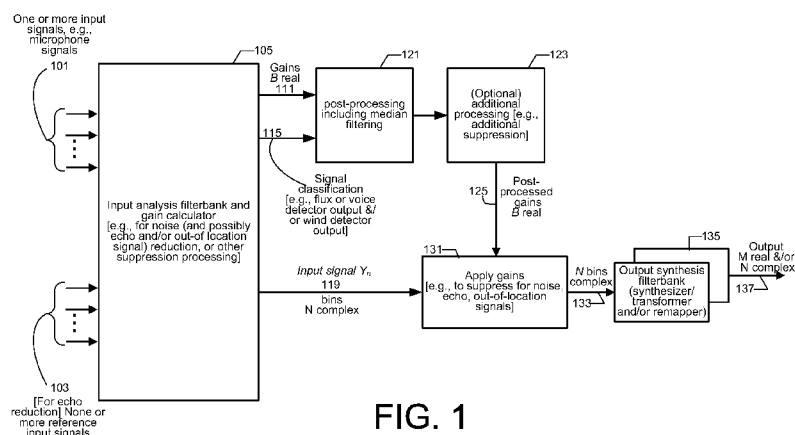
**Declarations under Rule 4.17:**

- as to the identity of the inventor (Rule 4.17(i))
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))
- of inventorship (Rule 4.17(iv))

**Published:**

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: POST-PROCESSING INCLUDING MEDIAN FILTERING OF NOISE SUPPRESSION GAINS

**FIG. 1**

(57) **Abstract:** A method of post-processing banded gains for applying to an audio signal, an apparatus to banded post-processed gains, and a tangible computer-readable storage medium comprising instructions that when executed carry out the method. The banded gains are determined by input processing one or more input audio signals. The method includes post-processing the banded gains to generate banded post-processed gains, generating a particular post-processed gain for a particular frequency band, including median filtering using gain values from gain values for frequency bands adjacent to the particular frequency band. One or more properties of the post-processing depends on classification of the one or more input audio signals.



## POST-PROCESSING INCLUDING MEDIAN FILTERING OF NOISE SUPPRESSION GAINS

### RELATED PATENT APPLICATIONS

**[0001]** The present application claims priority of United States Provisional Application  
5 No. 61/441,611 filed 10 February 2011, the contents of which are incorporated herein by  
reference in their entirety.

**[0002]** The present application is related to concurrently filed International Application  
No. PCT/US\_\_\_/\_\_\_\_\_, titled COMBINED SUPPRESSION OF NOISE, ECHO, and  
OUT-OF-LOCATION SIGNALS, filed with Docket/Ref. No. D09110AWO01, that also  
10 claims priority of United States Provisional Application No. 61/441,611 filed 10 February  
2011. The contents of such International Application No. PCT/US\_\_\_/\_\_\_\_\_ are  
incorporated herein by reference in their entirety.

**[0003]** The present application is related to the following U.S. provisional patent  
applications, each filed 10 February 2011:

15 **[0004]** • U.S. Provisional Patent Application No. 61/441,396, titled “VECTOR NOISE  
CANCELLATION” to inventor Jon C. Taenzer, Docket: 60175-0060, Client  
Ref. No. A09070USP1.

**[0005]** • U.S. Provisional Patent Application No. 61/441,397, titled “VECTOR NOISE  
CANCELLATION” to inventors Jon C. Taenzer and Steven H. Puthuff,  
20 Docket: 60175-0087, Client Ref. No. A09071USP1.

**[0006]** • U.S. Provisional Patent Application No. 61/441,528, titled “MULTI-  
CHANNEL WIND NOISE SUPPRESSION SYSTEM AND METHOD” to  
inventor Jon C. Taenzer, Docket: 435372-000052, Client Ref. No.  
D10025USP1.

25 **[0007]** • U.S. Provisional Patent Application No. 61/441,551, titled “SYSTEM AND  
METHOD FOR WIND DETECTION AND SUPPRESSION” to inventors  
Glenn N. Dickins and Leif Jonas Samuelsson, Docket: 435372-000053, Client  
Ref. No.: D10089USP1, such Provisional Patent Application No. 61/441,551  
being referred to as the “Wind Detection/Suppression Application” herein.

- [0008]**     • U.S. Provisional Patent Application No. 61/441,633, titled “SPATIAL ADAPTATION FOR MULTI-MICROPHONE SOUND CAPTURE” to inventor Leif Jonas Samuelsson, Docket: 435372-000054, Client Ref. No. D11008USP1.

## 5        FIELD OF THE INVENTION

- [0009]**     The present disclosure relates generally to signal processing, in particular of audio signals.

## BACKGROUND

- [0010]**     Acoustic signal processing is applicable today to improve the quality of sound signals such as from microphones. As one example, many devices such as handsets operate in the presence of sources of echoes, e.g., loudspeakers. Furthermore, signals from microphones may occur in a noisy environment, e.g., in a car or in the presence of other noise. Furthermore, there may be sounds from interfering locations, e.g., out-of-location conversation by others, or out-of-location interference, wind, etc. Acoustic signal processing is therefore an important area for invention.

- [0011]**     Processing systems are known for one or more of suppressing noise, suppressing echo, and adding spatial selectivity. An acoustic noise reduction system typically includes a noise estimator and a gain calculation module to determine suppression probability indicators, e.g., as a set of noise reduction gains that are determined, for example, on a set of frequency bands, and applied to the (noisy) input audio signal after transformation to the frequency domain and banding to the set of frequency bands to attenuate noise components. The acoustic noise reduction system may include one microphone input, or a plurality of microphone inputs and downmixing, e.g., beamforming to generate one input audio signal. The acoustic noise reduction system may further include echo reduction, and may further include out-of-location signal reduction.

- [0012]**     Musical noise is known to exist, and might occur because of short term mistakes over time made on the gain in some of the bands. Such gains-in-error can be considered statistical outliers, that is, values of the gain that across a group of bands statistically lie outside an expected range, so appear “isolated.”

[0013] Such statistical outliers might occur in other types of processing in which an input audio signal is transformed and banded. Such other types of processing include perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization that takes into account the variation in the perception of audio depending on the reproduction level of the audio signal. See, for example, International Application PCT/US2004/016964, published as WO 2004111994. It is possible that the gains determined for each band for leveling and/or dynamic equalization include statistical outliers, e.g., isolated values, and such outliers might cause artifacts such as musical noise.

10 [0014] Median filtering the gains, e.g., noise reduction gains, or leveling and/or dynamic equalization gains across frequency bands can reduce musical noise artifacts.

[0015] Gain values may vary significantly across frequencies, and in such a situation, running a relatively wide median filter along frequency bands has the risk of disrupting the continuity of temporal envelope, which is the inherent property for many signals and is crucial to perception as well. Whilst offering greater immunity to the outliers, a longer median filter can reduce the spectral selectivity of the processing, and potentially introduce greater discontinuities or jumps in the gain values across frequency and time.

[0016] The approaches described in this BACKGROUND section are approaches that could be pursued, but not necessarily approaches that have been previously conceived or pursued. Therefore, unless otherwise indicated, it should not be assumed that any of the approaches described in this section qualify as prior art merely by virtue of their inclusion in this section. Similarly, issues identified with respect to one or more approaches should not assume to have been recognized in any prior art on the basis of this section, unless otherwise indicated.

## 25 BRIEF DESCRIPTION OF THE DRAWINGS

[0017] FIG. 1 shows one example of processing of a set of one or more input audio signals, e.g., microphone signals 101 from differently located microphones, including an embodiment of the present invention.

[0018] FIG. 2 shows diagrammatically sets of raw banded gains and the frequency coverage of one embodiment of a median filter of the present invention.

[0019] FIG. 3A shows a simplified block diagram of a post-processor that includes a median filter according to an embodiment of the present invention.

[0020] FIG. 3B shows a simplified flowchart of a method of post-processing that includes median filtering according to an embodiment of the present invention.

5 [0021] FIG. 4 shows one example of an apparatus embodiment configured to determine a set of post-processed gains for suppression of noise, and in some versions, simultaneous echo suppression, and in some versions, simultaneous suppression of out-of-location signals.

[0022] FIG. 5 shows one example of an apparatus embodiment in more detail.

10 [0023] FIG. 6 shows an example embodiment of a gain calculation element that includes a spatially sensitive voice activity detector and a wind activity detector.

[0024] FIG. 7 shows a flowchart of an embodiment of a method of operating a processing apparatus to suppress noise and out-of-location signals and, in some embodiments, echoes.

15 [0025] FIG. 8 shows a simplified block diagram of a processing apparatus embodiment for processing one or more audio inputs to determine a set of raw gains, to post-process the raw gains including median filtering the determined raw gains, and to generate audio output that has been modified by application of the post-processed gains.

## DESCRIPTION OF EXAMPLE EMBODIMENTS

### 20 *Overview*

[0026] Embodiments of the present invention include a method, an apparatus, and logic encoded in one or more computer-readable tangible media to carry out the method.

[0027] One embodiment includes a method of applying post-processing to raw banded gains to improve the raw banded gains for applying to one or more input audio signals.

25 The raw banded gains at a plurality of frequency bands comprising one or more frequency bins are determined by input processing the one or more input audio signals. The raw banded gains are to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization. The method

comprises applying post-processing to the raw banded gains to generate banded post-processed gains. Generating of a particular post-processed gain for a particular frequency band includes at least median filtering using gain values for frequency bands adjacent to the particular frequency band. The post-processing is according to one or more properties, including an end condition and a width for the median filtering. At least one property of the post-processing depends on signal classification of the one or more input audio signals.

**[0028]** One embodiment includes a method of processing one or more input audio signals. The method includes input processing one or more input audio signals to determine raw banded gains for applying to an audio signal, the raw banded gains being at a plurality of frequency bands comprising one or more frequency bins. The raw banded gains are to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization. The method further comprises applying post-processing to the raw banded gains to generate banded post-processed gains. Generating of a particular post-processed gain for a particular frequency band includes at least median filtering using gain values for frequency bands adjacent to the particular frequency band. The post-processing is according to one or more properties, including an end condition and a width for the median filtering. At least one property of the post-processing depends on signal classification of the one or more input audio signals.

**[0029]** One embodiment includes an apparatus to post-process raw banded gains for applying to one or more input audio signals. The raw banded gains at a plurality of frequency bands comprising one or more frequency bins are determined by input processing the one or more input audio signals. The raw banded gains are to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization. The apparatus comprises a post-processor accepting the raw banded gains and applying post-processing to the raw banded gains to generate banded post-processed gains to apply to the one or more input signals. The post-processor includes a median filter to carry out median filtering of the raw banded gains. Generating by the post-processor of a particular post-processed gain for a particular frequency band

includes the median filtering using gain values for frequency bands adjacent to the particular frequency band. The post-processing is according to one or more properties, including an end condition and a width for the median filtering. At least one property of the post-processing depends on signal classification of the one or more input audio signals.

**[0030]** One embodiment includes an apparatus to process one or more input audio signals. The apparatus comprises an input processor accepting the one or more input audio signals and input processing the one or more input audio signals to generate raw banded gains at a plurality of frequency bands comprising one or more frequency bins. The raw banded gains are to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization. The apparatus further comprises a post-processor accepting the raw banded gains and applying post-processing to the raw banded gains to generate banded post-processed gains to apply to the one or more input signals. The post-processor includes a median filter to carry out median filtering of the raw banded gains. Generating by the post-processor of a particular post-processed gain for a particular frequency band includes the median filtering using gain values for frequency bands adjacent to the particular frequency band. The post-processing is according to one or more properties, including an end condition and a width for the median filtering. At least one property of the post-processing depends on signal classification of the one or more input audio signals.

**[0031]** One embodiment includes a system for post-processing raw banded gains to generate banded post-processed gains for applying to an audio signal. The system comprises means for post-processing raw banded gains to generate banded post-processed gains, the raw banded gains determined by a means for input processing one or more input audio signals to generate the raw banded gains at a plurality of frequency bands comprising one or more frequency bins. The raw banded gains are to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization. The banded post-processed gains are for applying to the one or more audio signals. Generating a particular post-processed gain for a particular frequency band includes at least median filtering using gain values for frequency bands

adjacent to the particular frequency band. The post-processing is according to one or more properties, including an end condition and a width for the median filtering. At least one property of the post-processing depends on signal classification of the one or more input audio signals.

5   **[0032]**     One embodiment includes a system for processing one or more input audio signals. The system comprises means for input processing the one or more input audio signals to generate raw banded gains at a plurality of frequency bands comprising one or more frequency bins. The raw banded gains are to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based  
10   leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization. The system further comprises means for post-processing raw banded gains to generate banded post-processed gains to apply to the one or more input audio signals to carry out the one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based  
15   dynamic range control, and perceptual domain-based dynamic equalization. The generating by the means for post-processing of a particular post-processed gain for a particular frequency band includes at least median filtering using gain values for frequency bands adjacent to the particular frequency band. The post-processing applied by the means for post-processing is according to one or more properties, including an end  
20   condition and a width for the median filtering. At least one property of the post-processing depends on signal classification of the one or more input audio signals.

**[0033]**     In some embodiments, the post-processing includes at least one of frequency-band-to-frequency-band smoothing and smoothing across time.

**[0034]**     In some embodiments, one of or both the width and the end conditions of the  
25   median filtering depend on signal classification of the one or more input audio signals. In some embodiments, the classification includes whether the input audio signals are likely or not to be voice. In some embodiments, the classification includes whether the input audio signals are likely or not to be wind.

**[0035]**     In some embodiments, the frequency bands are on a perceptual or logarithmic  
30   scale.

[0036] In some embodiments, the raw banded gains determined from one or more input audio signals are for reducing noise. In some embodiments, the raw banded gains are determined from more than one input audio signal and are for reducing noise and out-of-location signals. In some embodiments, the raw banded gains are determined from one or more input audio signals and one or more reference signals, and are for reducing noise and echoes.

[0037] One embodiment includes a tangible computer-readable storage medium comprising instructions that when executed by one or more processors of a processing system cause processing hardware to carry out a method of post-processing raw banded gains for applying to an audio signal as described herein.

[0038] One embodiment includes program logic that when executed by at least one processor causes carrying out a method as described herein.

[0039] Particular embodiments may provide all, some, or none of these aspects, features, or advantages. Particular embodiments may provide one or more other aspects, features, or advantages, one or more of which may be readily apparent to a person skilled in the art from the figures, descriptions, and claims herein.

### *Some example embodiments*

[0040] One aspect of the invention is processing of one or more input audio signals, including input processing to generate raw gains for noise reduction, or for other forms of input signal improvement. The processing includes applying post-processing to the raw gains, including median filtering of raw gains for gain smoothing. Another aspect of the invention is the post-processing that includes the median filtering of raw gains determined by input processing, e.g., for noise reduction or for other input processing. A median filter replaces a particular raw gain value with the median of a predefined number of raw gain values, e.g., by the median of the particular raw gain value and a predefined set of neighboring raw gain values. A median filter has one or more properties, e.g., the number of values of which the median is determined, and the end conditions. At least one of the properties may be data dependent. Therefore, in some examples described herein, there may be a first median filter for one type of data, e.g., data likely to be noise, and a different median filter for another type of data, e.g., data likely to be voice.

[0041] FIG. 1 shows one example of processing of a set of one or more input audio signals, e.g., microphone signals 101 from differently located microphones, including an embodiment of the present invention. The processing is by time frames of a number, e.g.,  $M$  samples. In a simple embodiment, there is only one input, e.g., one microphone, and in  
5 another embodiment, there is a plurality, denoted  $P$  of inputs, e.g., microphone signals 101. An input processor 105 accepts sampled input audio signal(s) 101 and forms a banded instantaneous frequency domain amplitude metric 119 of the input audio signal(s) 101 for a plurality  $B$  of frequency bands. In some embodiments in which there is more than one input audio signal, the metric 119 is mixed-down from the input audio signal.  
10 The amplitude metric represents the spectral content. In many of the embodiments described herein, the spectral content is in terms of the power spectrum. However, the invention is not limited to processing power spectral values. Rather, any spectral amplitude dependent metric can be used. For example, if the amplitude spectrum is used directly, such spectral content is sometimes referred to as spectral envelope. Thus, the  
15 phrase “power (or other amplitude metric) spectrum” is sometimes used in the description.

[0042] In one noise reduction embodiment, the input processor 105 determines a set of raw banded gains 111 to apply to the instantaneous amplitude metric 119. In one embodiment the input processing further includes determining a signal classification of the input audio signal(s), e.g., an indication of whether the input audio signal(s) is/are  
20 likely to be voice or not as determined by a voice activity detector (VAD), and/or an indication of whether the input audio signal(s) is/are likely to be wind or not as determined by a wind activity detector (WAD), and/or an indication that the signal energy is rapidly changing as indicated, e.g., by the spectral flux exceeding a threshold.

[0043] A feature of embodiments of the present invention includes applying post-  
25 processing to the raw gains to improve the quality of the output. In one embodiment the post-processing includes median filtering of the raw gains determined by the input processing. A median filter considers a set of raw gains and outputs the gain that is the median of the set of raw gains. A set of  $B$  raw gains is determined every frame, so that there is a time sequence of sets of  $B$  raw gains over  $B$  frequency bands. In one  
30 embodiment, the median filter extends across frequency.

[0044] FIG. 2 shows diagrammatically sets of raw banded gains, one set for each of the present time, one frame back, two frames back, three frames back, etc., and further shows

the coverage of an example median filter that includes five raw gain values centered around a frequency band  $b_c$  in the present frame. By filter width we mean the width of the filter in the frequency band domain.

[0045] Returning to FIG. 1, the post-processing produces a set of post-processed gains  
5 125 that are applied to the instantaneous power (or other amplitude metric) 119 to produce output, e.g., as a plurality of processed frequency bins 133. An output synthesis filterbank 135 (or for subsequent coding, a transformer/remapper) converts these frequency bins to desired output 137.

[0046] Input processing element 105 includes an input analysis filterbank, and a raw gain  
10 calculator. The input analysis filterbank, for the case of one input audio signal 101, includes a transformer to transform the samples of a frame into frequency bins, and a banding element to form frequency bands, most of which include a plurality of frequency bins. The input analysis filterbank, for the case of a plurality of input audio signals 101,  
15 includes a transformer to transform the samples of a frame of each of the input audio signals into frequency bins, a downmixer, e.g., a beamformer to downmix the plurality into a single signal, and a banding element to form frequency bands, most of which include a plurality of frequency bins.

[0047] In one embodiment, the transformer implements short time Fourier transform (STFT). For computational efficiency, the transformer uses a discrete finite length Fourier  
20 transform (DFT) implemented by a fast Fourier transform (FFT). Other embodiments use different transforms.

[0048] In one embodiment, the  $B$  bands are at frequencies whose spacing is monotonically non-decreasing. A reasonable number, e.g., 90% of the frequency bands include contribution from more than one frequency bin, and in particular embodiments,  
25 each frequency band includes contribution from two or more frequency bins. In some embodiments, the bands are monotonically increasing in a log-like manner. In some embodiments, the bands are on a psycho-acoustic scale, that is, the frequency bands are spaced with a scaling related to psycho-acoustic critical spacing, such banding called “perceptually-banding” herein. In particular embodiments, the band spacing is around 1  
30 ERB or 0.5 Bark, or equivalent bands with frequency separation at around 10% of the

center frequency. A reasonable range of frequency spacing is from 5–20% or approximately 0.5–2 ERB.

[0049] In some embodiments in which the input processing includes noise reduction, the input processing, a plurality of input audio signals are accepted by the input processor, and the input processing includes reducing out-of location signals. An example of input processing that includes reducing out-of location signals is described in concurrently filed International Application No. PCT/US\_\_\_/\_\_\_\_\_, titled COMBINED SUPPRESSION OF NOISE, ECHO, and OUT OF-LOCATION SIGNALS that also claims priority of U.S. Provisional Application No. 61/441,611 filed 10 February 2011 to inventors Dickins et al. titled “COMBINED SUPPRESSION OF NOISE, ECHO, AND OUT-OF-LOCATION SIGNALS,” the contents of both which are incorporated herein by reference. The resulting raw banded gains achieve simultaneous echo reduction and noise reduction.

[0050] In some embodiments in which the input processing includes noise reduction, the input processing also includes echo reduction. One example of input processing that includes echo reduction is described in concurrently filed International Application No. PCT/US. For those embodiments in which the input processing includes echo reduction, one or more reference signals also are included and used to obtain an estimate of some property of the echo, e.g., of the power (or other amplitude metric) spectrum of the echo. The resulting raw banded gains achieve simultaneous echo reduction and noise reduction.

[0051] In some embodiments that include noise reduction and echo reduction, the post-processed gains are accepted by an element 123 that modifies the gains to include additional echo suppression. The result is a set of post-processed gains 125 that are applied to the input signal or signals. e.g., that are used to process, in the frequency domain, as frequency bins, the input audio signal if there is one input, or a downmix of the input audio signals if there is a plurality of input audio signals, e.g., from differently located microphones.

[0052] Gain application module 131 accepts the banded post-processed gains 125 and applies such gains to the input audio signal or signals. In one embodiment, the banded gains are interpolated and applied to the frequency bin data of the input audio signal (if one) or the downmixed input audio signal (if there is more than one input audio signal), denoted  $Y_n$ ,  $n=0, 1, \dots, N-1$ , where  $N$  is the number of frequency bins.  $Y_n$ ,  $n=0, 1, \dots, N-$

1 are the frequency bins of a frame of input audio signal samples  $Y_m$ ,  $m=1, \dots, M$ . The processed data 133 may then be converted back to the sample domain by an output synthesis filterbank 135 to produce a frame of  $M$  signal samples 137. In some embodiments, in addition or instead, the signal 133 is subject to transformation or  
5 remapping, e.g., to a form ready for coding according to some coding method.

**[0053]** An example embodiment of a system similar to that of PCT/US\_\_\_/\_\_\_\_\_ that includes input processing to reduce noise (and possibly echo and out of location signals) is described in more detail below.

**[0054]** The invention, of course, is not limited to the input processing and gain calculation  
10 described in International Application No. PCT/US\_\_\_\_\_, U.S. 61/441,611, or even to noise reduction.

**[0055]** While in one embodiment the input processing is to reduce noise (and possibly echo and out of location signals), in other embodiments, the input processing may be, additionally or primarily, to carry out one or more of perceptual domain-based leveling,  
15 perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization that take into account the variation in the perception of audio depending on the reproduction level of the audio signal, as described, for example, in commonly owned WO 2004111994. The raw banded gains calculated per WO 2004111994 are post-processed, including median filtering, to determine post-processed gains 125 to apply to  
20 the (transformed) input.

### *Post-processing including median filtering*

**[0056]** FIG. 3A shows a simplified block diagram of a post-processor 121 that includes a median filter 305 according to an embodiment of the present invention. The post-processor 121 accepts raw gains 111 and in embodiments in which the post-processing  
25 changes according to signal classification, one or more signal classification indicators 115, e.g., the outputs of one or more of a VAD and a WAD. While not included in all embodiments, some embodiments of the post-processor include a minimum gain processor 303 to ensure that the gains do not fall below a predefined, possibly frequency-dependent value. Again while not included in all embodiments, some embodiments of the  
30 post-processor include a smoothing filter 307 that processes the gains after median filtering to smooth frequency-band-to-frequency-band variations, and/or to smooth time

variations. FIG. 3B shows a simplified flowchart of a method of post-processing 310 that includes in 311 accepting raw gains, and in embodiments in which the post-processing changes according to signal classification, one or more signal classification indicators 115. The post-processing includes median filtering 315 according to embodiments of the present invention. The inventors have found that median filtering is a powerful nonlinear smoothing technique, which works well for eliminating undesired outliers when compared with only using a smoothing method. Some embodiments include in step 313 ensuring that the gains do not fall below a predefined minimum, which may be frequency band dependent. Some embodiments further include, in step 317, band-to-band and/or time smoothing, e.g., linear smoothing using, e.g., a weighted moving average.

**[0057]** Thus, in some embodiment of the present invention, a median filter 315 of raw banded gain values is characterized by: 1) the number of raw banded gains to include to determine the median value, 2) the frequency band positions of the raw banded gains that are included; 3) the edge conditions, i.e., the conditions used to extend the raw banded gains to allow calculation of the median at the edges of time and frequency band; and 4) how the characterization of the median filter is affected by the signal classification, e.g., one or more of the presence of voice and the presence of wind.

**[0058]** Some embodiments include a mechanism to control one or more of the median filtering characteristics over frequency and/or time based on signal classification. For example, in one embodiment that includes voice activity detection, one or more of the median filtering characteristics vary in accordance to whether the input is ascertained by a VAD to be voice or not. In one embodiment that includes wind activity detection, one or more of the median filtering characteristics vary in accordance to whether the input is ascertained by a WAD to be wind or not.

**[0059]** Examples of different edge conditions include (a) extrapolating of interior values for the edges; (b) using the minimum gain value to extend the raw banded gains at the edges, (c) using a zero gain value to extend the raw banded gains at the edges (d) duplicating the central filter position value to extend the raw banded gains at the edges, and (e) using a maximum gain value to extend the raw banded gains at the edges.

**Additional post-processing**

[0060] While not included in all embodiments, in some embodiments the post-processor 121 includes a minimum gain processor 303 that carries out step 313 to ensure the gains do not fall below a predefined minimum gain value. In some embodiments, the minimum gain processor ensures minimum values in a frequency-band dependent manner. In some  
 5 embodiments, the manner of prevention minimum is dependent on the activity classification 115, e.g., whether voice or not.

[0061] In one embodiment, denoting the calculated raw gains from the input processing by  $Gain'_{b,S}$ , some alternatives for the gains denoted  $Gain'_{b,RAW}$  after minimum processor  
 10 are

$$[0062] \quad Gain'_{b,RAW} = Gain'_{b,MIN} + (1 - Gain'_{b,MIN}) \cdot Gain'_{b,S}$$

$$Gain'_{b,RAW} = Gain'_{b,MIN} + Gain'_{b,S}$$

$$Gain'_{b,RAW} = \begin{cases} Gain'_{b,MIN} & Gain'_{b,S} < Gain'_{b,MIN} \\ Gain'_{b,S} & \text{otherwise} \end{cases}$$

[0063] As one example, in some embodiments of post-processor 121 and step 310, the  
 15 range of the maximum suppression depth or minimum gain may range from  $-80\text{dB}$  to  $-5\text{dB}$  and be frequency dependent. In one embodiment the suppression depth was around  $-20\text{dB}$  at low frequencies below  $200\text{Hz}$ , varying to be around  $-10\text{dB}$  at  $1\text{kHz}$  and relaxing to be only  $-6\text{dB}$  at the upper voice frequencies around  $4\text{kHz}$ . Furthermore, in one embodiment, if a VAD determines the signal to be voice,  $Gain'_{b,MIN}$  is increased, e.g., in  
 20 a frequency-band dependent way (or in another embodiment, by the same amount for each band  $b$ ). In one embodiment, the amount of increase in the minimum is larger in the mid-frequency bands, e.g., bands between  $500\text{ Hz}$  to  $2\text{ kHz}$ .

[0064] Furthermore, while not included in all embodiments, in some embodiments the post-processor 121 includes a smoothing filter 307, e.g., a linear smoothing filter that  
 25 carries out one or both of frequency band-to-band smoothing and time smoothing. In some embodiments, such smoothing is varied according to signal classification 115.

[0065] One embodiment of smoothing 317 uses a weighted moving average with a fixed kernel. One example uses a binomial approximation of a Gaussian weighting kernel for

the weighted moving average. As one example, a 5-point binomial smoother has a kernel

$\frac{1}{16}[1 \ 4 \ 6 \ 4 \ 1]$ . In practice, of course, the factor 1/16 may be left out, with scaling

carried out in one point or another as needed. As another example, a 3-point binomial

smoother has a kernel  $\frac{1}{4}[1 \ 2 \ 1]$ . Many other weighted moving average filters are

5 known, and any such filter can suitably be modified to be used for the band-to-band smoothing of the gain.

**[0066]** In one embodiment, the band-to-band smoothing is controlled by the signal classification. In one embodiment, a VAD, e.g., a spatially-selective VAD is included, and if the VAD determines there is voice, the degree of smoothing is increased when noise is  
10 detected. In one example embodiment, 5-point band-to-band weighted average smoothing is carried out in the case the VAD indicates voice is detected, else, when the VAD determines there is no voice, no smoothing is carried out.

**[0067]** In some embodiments, time smoothing of the gains also is included. In some embodiments, the gain of each of the  $B$  bands is smoothed by a first order smoothing  
15 filter:

**[0068]**  $Gain_{b,Smoothed} = \alpha_b Gain_b + (1 - \alpha_b) Gain_{b,Smoothed_{Prev}}$

where  $Gain_b$  is the current time-frame gain,  $Gain_{b,Smoothed}$  is the time-smoothed gain, and

$Gain_{b,Smoothed_{Prev}}$  is  $Gain_{b,Smoothed}$  from the previous  $M$ -sample frame.  $\alpha_b$  is a time

constant which may be frequency band dependent and is typically in the range of 20 to

20 500ms. In one embodiment a value of 50ms was used. In one embodiment, the amount of time smoothing is controlled by the signal classification of the current frame. In a particular embodiment that includes first order time smoothing of the gains, the signal classification of the current frame is used to control the values of first order time constants used to filter the gains over time in each band. In the case a VAD is included, one  
25 embodiment stops time smoothing in the case voice is detected.

**[0069]** The inventors found it is important that aggressive smoothing be discontinued at the onset of voice. Thus it is preferable that the parameters of post-processing are controlled by the immediate signal classifier (VAD, WAD) value that has low latency and is able to achieve a rapid transition of the post-processing from noise into voice (or other

desired signal) mode. The speed with which more aggressive post-processing is reinstated after detection of voice, i.e., at the trail out, has been found to be less important, as it affects intelligibility of voice to a lesser extent.

#### **An example of voice activity control**

5   **[0070]**     In one embodiment, the band-to-band median filtering is controlled by the signal classification. In one embodiment, a VAD is included, and if the VAD determines it is likely that there is no voice, a 7 point T-shaped median filter with 5-point band-to-band and 3-point time median filtering is carried out, with edge processing including extending minimum gain values or a zero value at the edges to compute the median value. If the  
10   VAD determines it is likely that voice is present, in a first version, a 5-point T-shaped time-frequency median filtering is carried out with three frequency bands in the current time frame, and using two previous time frames, and in a second embodiment, a three point memoryless frequency-band only median filter, with the edge values extrapolated at the edges to calculate the median, is used. In one such set of embodiments, the median  
15   value is the median value, such that the median filter is a median filter.

#### **An example of wind activity control**

**[0071]**     One feature of the present invention is that the post-processing, e.g., the median filtering depends on the classification of the signal, and one such classification, in some embodiments, is whether there is wind or not. In some embodiments, a WAD is included,  
20   and if the WAD determines there is no wind, and a VAD indicates there is no voice, fewer raw gain values are included in the median filter. When WAD and a VAD is included, if the WAD determines there is likely not to be wind and the VAD determines voice is likely, the median filtering should be shorter, e.g., by using 3-point band-to-band median filter, with extrapolating the edge values applied at the edges. If the WAD indicated wind  
25   is unlikely, and the VAD indicates voice is also unlikely, more median filtering can be used, e.g., 5-point band-to-band median filtering is carried out, with edge processing including extending minimum gain values or a zero value at the edges to compute the median value. If the WAD indicated wind is likely, and the VAD indicates voice is unlikely, even more median filtering can be used, e.g., a 7-point band-to-band median  
30   filtering can be carried out, with edge processing including extending minimum gain values or a zero value at the edges to compute the median value.

### *An example acoustic noise reduction system*

[0072] An acoustic noise reduction system typically includes a noise estimator and a raw gain calculation module to determine a set of noise reduction gains that are determined, for example, on a set of frequency bands, and applied to the (noisy) input audio signal after transformation to the frequency domain and banding to the set of frequency bands to attenuate noise components. The acoustic noise reduction system may include one microphone, or a plurality of inputs from differently located microphones and downmixing, e.g., beamforming to generate one input audio signal. The acoustic noise reduction system may further include echo reduction, and may further include out-of-location signal reduction.

[0073] FIG. 4 shows one example of an apparatus configured to determine a set of post-processed gains for suppression of noise, and in some versions, simultaneous echo suppression, and in some versions, simultaneous suppression of out-of-location signals. Such a system is described, e.g., in International Application PCT/US\_\_\_/\_\_\_\_\_ and in US 61/441,611. The inputs include a set of one or more input audio signals 101, e.g., signals from differently located microphones, each in sets of  $M$  samples per frame. When spatial information is included, there are two or more input audio signals, e.g., signals from spatially separated microphones. When echo suppression is included, one or more reference signals 103 are also accepted, e.g., in frames of  $M$  samples. These may be, for example, one or more signals from one or more loudspeakers, or, in another embodiment, the signal(s) that are used to drive the loudspeaker(s). A first input processing stage 403 determines a banded signal power (or other amplitude metric) spectrum 413 denoted  $P'_b$ , and a banded measure of the instantaneous power 417 denoted  $Y'_b$ . When more than one input audio signal is included, each of the spectrum 413 and instantaneous banded measure 417 is of the inputs after being mixed down by a downmixer, e.g., a beamformer. When echo suppression is included, the first input processing stage 403 also determines a banded power spectrum estimate of the echo 415, denoted  $E'_b$ , the determining being from a previously calculated power spectrum estimates of the echo using a filter with a set of adaptively determined filter coefficients. In those versions that include out-of-location signal suppression, the first input processing stage 403 also determines spatial features 419 in the form of banded location probability indicators 419 that are usable to spatially

separate a signal into the components originating from the desired location and those not from the desired direction.

**[0074]** The quantities from the first stage 403 are used in a second stage 405 that determines raw gains, and that post-processes the raw gains, including the median  
 5 filtering of embodiments of the present invention, to determine the banded post-processed gains 125. Embodiments of the second stage 405 include a noise power (or other amplitude metric) spectrum calculator 421 to determine a measure of the noise power (or other amplitude metric) spectrum, denoted  $E'_b$ , and a signal classifier 423 to determine a signal classification 115, e.g., one or more of a voice activity detector (VAD), a wind  
 10 activity detector, and a power flux calculator. FIG. 4 shows the signal classifier 423 including a VAD.

**[0075]** FIG. 5 shows one embodiment 500 of the elements of FIG. 4 in more detail, and includes, for the example embodiment of noise, echo, and out-of-location noise suppression, the suppressor 131 that applied the post-processed gains 125 and the output  
 15 synthesizer (or transformer or remapper) 135 to generate the output signal 137.

**[0076]** Comparing FIGS. 4 and 5, the first stage processor 403 of FIG. 4 includes elements 503, 505, 507, 509, 511, 513, 515, 517, 521, 523, 525, and 527 of FIG. 5. In more detail, the input(s) frame(s) 101 are transformed by inputs transformer(s) 503 to determine transformed input signal bins, the number of frequency bins denoted by  $N$ . In  
 20 the case of more than one input audio signal, these frequency domain signals are beamformed by a beamformer 507 to form input frequency bin data denoted  $Y_n$ ,  $n=1, \dots, N$ , and the input frequency bin data  $Y_n$  is banded by spectral banding element 509 into  $B$  spectral bands, in one embodiment, perceptually spaced spectral bands to produce the instantaneous banded measure of the power  $Y'_b$ ,  $b=1, \dots, B$ . In a version that includes out-  
 25 of-location suppression and more than one input audio signal, the frequency domain signals from the input transformers 503 are accepted by a banded spatial feature calculator to determine banded location probability indicators, each between 0 and 1. In a version that includes echo suppression, if there is more than one reference signal, say  $Q$  reference signals, the signals are combined by combiner 511, in one embodiment a summer, to  
 30 produce a combined reference input. An input transformer 513 and spectral bander 515 convert the reference into banded reference spectral content denoted  $X'_b$ ,  $b=1, \dots, B$  for

the  $B$  bands. An  $L$ -tap linear prediction filter 517 predicts the banded echo spectral content  $E'_b$ ,  $b=1, \dots, B$ , using  $L$  times  $B$  filter update coefficients 528. A signal spectral calculator 521 calculates a measure of the (mixed-down) power (or other amplitude metric) spectrum  $P'_b$ ,  $b=1, \dots, B$ . In some embodiments,  $Y'_b$  is used as a good-enough  
 5 approximation to  $P'_b$ .

**[0077]** The  $L B$  filter coefficients for filter 517 are determined by an adaptive filter updater 527 that uses the current banded echo spectral content  $E'_b$ , the measure of the (mixed-down) power (or other amplitude metric) spectrum  $P'_b$ , a banded noise power (or other amplitude metric) spectrum 524 denoted  $N'_b$ ,  $b=1, \dots, B$ , and determined by a noise  
 10 calculator 523 from the instantaneous power  $Y'_b$  and a measure from the signal spectral calculator 521. The updating is triggered by a voice activity signal denoted  $S$  as determined by a voice activity detector (VAD) 525 using  $P'_b$  (or  $Y'_b$ ),  $N'_b$ , and  $E'_b$ . When  $S$  exceeds a threshold, the signal is assumed to be voice. The VAD derived in the echo update voice-activity detector 525 and filter updater 527 serves the specific purpose of  
 15 controlling the adaptation of the echo prediction. A VAD or detector with this purpose is often referred to as a double talk detector. In one embodiment, the echo filter coefficient updating of updater 527 is gated, with updating occurring when the expected echo is significant compared to the expected noise and current input power, as determined by the VAD 525 and indicated by a low value of local signal activity  $S$ .

20 **[0078]** Details of how the elements the first stage 403 per FIGS. 4 and 5 operate in some embodiments are as follows. In one embodiment, the input transformers 503, 511 determine the short time Fourier transform (STFT). In another embodiment, the following transform and inverse pair is used for the forward transform in elements 503 and 511, and in output synthesis element 135.

20

$$X_{2n} = \frac{1}{\sqrt{N}} \sum_{n'=0}^{N-1} e^{\frac{-i\pi n'}{2N}} (u_{n'} x_{n'} - i u_{N+n'} x_{N+n'}) e^{\frac{-i2\pi n n'}{N}} \quad n = 0 \dots N/2 - 1$$

$$X_{2n+1} = \frac{1}{\sqrt{N}} \sum_{n'=0}^{N-1} e^{\frac{i\pi n'}{2N}} (u_{n'} x_{n'} + i u_{N+n'} x_{N+n'}) e^{\frac{-i2\pi n n'}{N}} \quad n = 0 \dots N/2 - 1$$

**[0079]**

$$y_n = v_n \operatorname{real} \left[ \frac{1}{\sqrt{N}} e^{\frac{i\pi n}{4N}} \left( \sum_{n'=0}^{N/2-1} X_{n'} e^{\frac{i4\pi n n'}{N}} + \sum_{n'=N/2}^{N-1} X_{N-n'-1} e^{\frac{i4\pi n n'}{N}} \right) \right] \quad n = 0 \dots N-1 \quad \text{wh}$$

$$y_{N+n} = -v_{N+n} \operatorname{imag} \left[ \frac{1}{\sqrt{N}} e^{\frac{i\pi n}{4N}} \left( \sum_{n'=0}^{N/2-1} X_{n'} e^{\frac{i4\pi n n'}{N}} + \sum_{n'=N/2}^{N-1} X_{N-n'-1} e^{\frac{i4\pi n n'}{N}} \right) \right] \quad n = 0 \dots N-1$$

ere  $i^2 = -1$ ,  $u_n$  and  $v_n$  are appropriate window functions,  $x_n$  represents the last  $2N$  input samples with  $x_{N-1}$  representing the most recent sample,  $X_n$  represents the  $N$  complex-valued frequency bins in increasing frequency order. The inverse transform or synthesis is represented in the last two equation lines.  $y_n$  represents the  $2N$  output samples that result from the individual inverse transform prior to overlapping, adding and discarding as appropriate for the designed windows. It should be noted, that this transform has an efficient implementation as a block multiply and FFT. Note that the use of  $x_n$  and  $X_n$  in the above expressions of transform is for convenience. In other parts of this disclosure,  $X_n$ ,  $n=0, \dots, N-1$ , denote the frequency bins of the signal representative of the reference signals, and  $Y_n$ ,  $n=0, \dots, N-1$ , denote the frequency bins of the mixed-down input audio signals.

**[0080]** In one embodiment, the window functions  $u_n$  and  $v_n$  for the above transform in one embodiment is the sinusoidal window family, of which one suggested embodiment is

$$u_n = v_n = \sin \left( \frac{n + \frac{1}{2}}{2N} \pi \right) \quad n = 0 \dots 2N - 1.$$

**[0082]** It should be apparent to one skilled in the art that the analysis and synthesis windows, also known as prototype filters, can be of length greater or smaller than the examples given herein.

**[0083]** While the invention works with any mixed-down signal, in some embodiments, the downmixer is a beamformer 507 designed to achieve some spatial selectivity towards the desired position. In one embodiment, the beamformer 507 is a linear time invariant

process, i.e., a passive beamformer defined in general by a set of complex-valued frequency-dependent gains for each input channel. For the example of a two-microphone array, with the desired sound source located broad side to the array, i.e., at the perpendicular bisector, one embodiment uses for beamformer 507 a passive beamformer 107 that determines the simple sum of the two input channels. In some versions, beamformer 507 weights the sets of inputs (as frequency bins) by a set of complex valued weights. In one embodiment, the beamforming weights of beamformer 107 are determined according to maximum-ratio combining (MRC). In another embodiment, the beamformer 507 uses weights determined using zero-forcing. Such methods are well known in the art.

**[0084]** The banding of spectral banding elements 509 and 514 can be described by

$$\text{[0085]} \quad Y'_b = W_b \sum_{n=0}^{N-1} w_{b,n} |Y_n|^2$$

where  $Y'_b$  is the banded instantaneous power of the mixed-down, e.g., beamformed signal,  $W_b$  is the normalization gain and  $w_{b,n}$  are elements from a banding matrix.

**[0086]** The signal spectral calculator 521 in one embodiment is described by a smoothing process

$$\text{[0087]} \quad P'_b = \alpha_{P,b} (Y'_b + Y'_{\min}) + (1 - \alpha_{P,b}) P'_{b\text{PREV}},$$

where  $P'_{b\text{PREV}}$  is a previously, e.g., the most recently determined signal power (or other frequency domain amplitude metric) estimate,  $\alpha_{P,b}$  is a time signal estimate time constant, and  $Y'_{\min}$  is an offset. A suitable range for the signal estimate time constant  $\alpha_{P,b}$  was found to be between 20 to 200 ms. In one embodiment, the offset  $Y'_{\min}$  is added to avoid a zero level power spectrum (or other amplitude metric spectrum) estimate.  $Y'_{\min}$  can be measured, or can be selected based on a priori knowledge.  $Y'_{\min}$ , for example, can be related to the threshold of hearing or the device noise threshold.

**[0088]** In one embodiment, the adaptive filter 517 includes determining the instantaneous echo power spectrum (or other amplitude metric spectrum), denoted  $T'_b$  for band  $b$  by using an  $L$  tap adaptive filter described by

$$[0089] \quad T'_b = \sum_{l=0}^{L-1} F_{b,l} X'_{b,l} ,$$

where the present frame is  $X'_b = X'_{b,0}$ , where  $X'_{b,0}, \dots, X'_{b,l}, \dots, X'_{b,L-1}$  are the  $L$  most recent frames of the (combined) banded reference signal  $X'_b$ , including the present frame  $X'_b = X'_{b,0}$ , and where the  $L$  filter coefficients for a given band  $b$  are denoted by  $F_{b,0}, \dots,$   
 5  $F_{b,l}, \dots, F_{b,L-1}$ , respectively.

[0090] One embodiment includes time smoothing of the instantaneous echo from echo prediction filter 517 to determine the echo spectral estimate  $E'_b$ . In one embodiment, a first order time smoothing filter is used as follows

$$[0091] \quad E'_b = T'_b \text{ for } T'_b \geq E'_{b\text{Prev}}, \text{ and}$$

$$10 \quad E'_b = \alpha_{E,b} T'_b + (1 - \alpha_{E,b}) E'_{b\text{Prev}} \text{ for } T'_b < E'_{b\text{Prev}}$$

where  $E'_{b\text{Prev}}$  is the previously determined echo spectral estimate, e.g., in the most recently, or other previously determined estimate, and  $\alpha_{E,b}$  is a first order smoothing time constant.

[0092] In one embodiment, the noise power spectrum calculator 523 uses a minimum  
 15 follower with exponential growth:

$$[0093] \quad N'_b = \min(P'_b, (1 + \alpha_{N,b}) N'_{b\text{Prev}}) \text{ when } E'_b \text{ is less than } N'_{b\text{Prev}}$$

$$N'_b = N'_{b\text{Prev}} \text{ otherwise,}$$

where  $\alpha_{N,b}$  is a parameter that specifies the rate over time at which the minimum  
 follower can increase to track any increase in the noise. In one embodiment, the criterion  
 20  $E'_b$  is less than  $N'_{b\text{Prev}}$  is if  $E'_b < N'_{b\text{Prev}}/2$ , i.e., in the case that the (smoothed) echo spectral estimate  $E'_b$  is less than the previous value of  $N'_b$  less 3dB, in which case the noise estimate follows the growth or current power. Otherwise,  $N'_b = N'_{b\text{Prev}}$ , i.e.,  $N'_b$  is held at the previous value of  $N'_b$ . The parameter  $\alpha_{N,b}$  is best expressed in terms of the rate over time at which minimum follower will track. That rate can be expressed in

dB/sec, which then provides a mechanism for determining the value of  $\alpha_{N,b}$ . The range is 1 to 30dB/sec. In one embodiment, a value of 20dB/sec is used.

**[0094]** In other embodiments, different approaches for noise estimation may be used.

Examples of such different approaches include but are not limited to alternate methods of determining a minimum over a window of signal observation, e.g., a window of 1 and 10 seconds. In addition or alternate to the minimum, such different approaches might also determine the mean and variance of the signal during times that it is classified as likely to be noise or that voice is unlikely.

**[0095]** In one embodiment, the one or more leak rate parameters of the minimum follower are controlled by the probability of voice being present as determined by voice activity detecting (VAD). In one embodiment, VAD element 525 determines an overall signal activity level denoted  $S$  as

$$\text{[0096]} \quad S = \sum_{b=1}^B \frac{\max(0, Y'_b - \beta_N N'_b - \beta_E E'_b)}{Y'_b + Y'_{sens}}$$

where  $\beta_N, \beta_E > 1$  are margins for noise and echo, respectively and  $Y'_{sens}$  is a settable sensitivity offset. These parameters may in general vary across the bands. In one embodiment, the values of  $\beta_N, \beta_E$  are between 1 and 4. In a particular embodiment,  $\beta_N, \beta_E$  are each 2.  $Y'_{sens}$  is set to be around expected microphone and system noise level, obtained by experiments on typical components. Alternatively, one can use the threshold of hearing to determine a value for  $Y_{sens}$ .

**[0097]** In one embodiment, the echo filter coefficient updating of updater 527 is gated, as follows. If the local signal activity level is low, e.g., below a pre-defined threshold  $S_{thresh}$ , i.e., if  $S < S_{thresh}$ , then the adaptive filter coefficients are updated as:

$$\text{[0098]} \quad F_{b,l} = F_{b,l} + \mu \frac{(\max(0, Y'_b - \gamma_N N'_b) - T'_b) X'_{b,l}}{\sum_{l'=0}^{L-1} (X'_{b,l'}^2 + X'_{sens}^2)} \text{ if } S < S_{thresh},$$

where  $\gamma_N$  is a tuning parameter tuned to ensure stability between the noise and echo estimate. A typical value for  $\gamma_N$  is 1.4 (+3dB). A range of values 1 to 4 can be used.  $\mu$  is a tuning parameter that affects the rate of convergence and stability of the echo estimate.

Values between 0 and 1 might be useful in different embodiments. In one embodiment,  $\mu = 0.1$  independent of the frame size  $M$ .  $X'_{sens}$  is set to avoid unstable adaptation for small reference signals. In one embodiment  $X'_{sens}$  is related to the threshold of hearing. The choice of value for  $S_{thresh}$  depends on the number of bands.  $S_{thresh}$  is between 1 and  $B$ , and  
 5 for one embodiment having 24 bands to 8kHz, a suitable range was found to be between 2 and 8, with a particular embodiment using a value of 4.

**[0099]** Embodiments of the present invention use spatial information in the form of one or more measures determined from one or more spatial features in a band  $b$  that are monotonic with the probability that the particular band  $b$  has such energy incident from a spatial region of interest. Such quantities are called spatial probability indicators. In one  
 10 embodiment, the one or more spatial probability indicators are functions of one or more banded weighted covariance matrices of the input audio signals. Given the output of the  $P$  input transforms  $X_{p,n}$ ,  $p=1, \dots, P$ , with  $N$  frequency bins,  $n=0, \dots, N-1$ , we construct a set of weighted covariance matrices to correspond by summing the product of the input  
 15 vector across the  $P$  inputs for bin  $n$  with its conjugate transpose, and weighting by a banding matrix  $\mathbf{W}_b$  with elements  $w_{b,n}$

$$\mathbf{R}'_b = \sum_{n=0}^{N-1} w_{b,n} [X_{1,n} \quad \dots \quad X_{P,n}]^H [X_{1,n} \quad \dots \quad X_{P,n}].$$

**[00101]** The  $w_{b,n}$  provide an indication of how each bin is weighted for contribution to the bands. In some embodiments, the one or more covariance matrices are smoothed over  
 20 time. In some embodiments, the banding matrix includes time dependent weighting for a weighted moving average, denoted as  $\mathbf{W}_{b,l}$  with elements  $w_{b,n,l}$ , where  $l$  represents the time frame, so that, over  $L$  time frames,

$$\mathbf{R}'_b = \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} w_{b,n,l} [X_{1,n} \quad \dots \quad X_{P,n}]^H [X_{1,n} \quad \dots \quad X_{P,n}].$$

**[00103]** In the case of two inputs,  $P = 2$ , define

[00104]  $\mathbf{R}'_b = \begin{bmatrix} R'_{b11} & R'_{b12} \\ R'_{b21} & R'_{b22} \end{bmatrix},$

so that each band covariance matrix  $\mathbf{R}'_b$  is a 2x2 Hermetian positive definite matrix with  $R'_{b21} = \overline{R'_{b12}}$ , where the overbar is used to indicate the complex conjugate.

[00105] Denote by the spatial feature “ratio” a quantity that is monotonic with the ratio of  
 5 the banded magnitudes  $\frac{R'_{b11}}{R'_{b22}}$ . In one embodiment, a log relationship is used:

[00106]  $Ratio'_b = 10 \log_{10} \frac{R'_{b11} + \sigma}{R'_{b22} + \sigma}$

where  $\sigma$  is a small offset added to avoid singularities.  $\sigma$  can be thought of as the  
 smallest expected value for  $R'_{b11}$ . In one embodiment, it is the determined, or estimated (a  
 priori) value of the noise power (or other frequency domain amplitude metric) in band  $b$   
 10 for the microphone and related electronics. That is, the minimum sensitivity of any  
 preprocessing used.

[00107] Denote by the spatial feature phase a quantity monotonic with  $\tan^{-1} R'_{b21}$ .

[00108]  $Phase'_b = \tan^{-1} R'_{b21}.$

[00109] Denote by the spatial feature “coherence” a quantity that is monotonic with  
 15  $\frac{R'_{b21}R'_{b12}}{R'_{b11}R'_{b22}}$ . In some embodiments, related measures of coherence could be used such as  
 $\frac{2R'_{b21}R'_{b12}}{R'_{b11}R'_{b11} + R'_{b22}R'_{b22}}$  or values related to the conditioning, rank or eigenvalue spread of  
 the covariance matrix. In one embodiment, the coherence feature is

[00110]  $Coherence'_b = \sqrt{\frac{R'_{b21}R'_{b12} + \sigma^2}{R'_{b11}R'_{b22} + \sigma^2}}.$

[00111] with offset  $\sigma$  as defined above.

20 [00112] One feature of some embodiments of the noise, echo and out-of-location signal  
 suppression is that, based on the a priori expected or current estimate of the desired signal  
 features—the target values, e.g., representing spatial location, gathered from statistical

data—each spatial feature in each band can be used to create a probability indicator for the feature for the band  $b$ .

[00113] In one embodiment, the distributions of the expected spatial features for the desired location are modeled as Gaussian distributions that present a robust way of capturing the region of interest for probability indicators derived from each spatial feature and band.

[00114] Three spatial probability indicators are related to these three spatial features, and are the ratio probability indicator, denoted  $RPI'_b$ , the phase probability indicator, denoted  $PPI'_b$ , and the coherence probability indicator, denoted  $CPI'_b$ , with

$$[00115] \quad RPI'_b = f_{R_b} (Ratio'_b - Ratio_{target_b}) = f_{R_b} (\Delta Ratio'_b),$$

where  $\Delta Ratio'_b = Ratio'_b - Ratio_{target_b}$  and  $Ratio_{target_b}$  is determined from either prior estimates or experiments on the equipment used, e.g., headsets, e.g., from data such as shown in FIG. 9A.

[00116] The function  $f_{R_b} (\Delta Ratio')$  is a smooth function. In one embodiment, the ratio probability indicator function is

$$[00117] \quad f_{R_b} (\Delta Ratio') = \exp \left[ -\frac{\Delta Ratio'_b}{Width_{Ratio,b}} \right]^2,$$

where  $Width_{Ratio,b}$  is a width tuning parameter expressed in log units, e.g., dB. The  $Width_{Ratio,b}$  is related to but does not need to be determined from actual data. It is set to cover the expected variation of the spatial feature in normal and noisy conditions, but also needs only be as narrow as is required in the context of the overall system to achieve the desired suppression.

[00118] For the phase probability indicator,

$$[00119] \quad PPI'_b = f_{P_b} (Phase'_b - Phase_{target_b}) = f_{P_b} (\Delta Phase'_b),$$

where  $\Delta Phase'_b = Phase'_b - Phase_{target_b}$  and  $Phase_{target_b}$  is determined from either prior estimates or experiments on the equipment used, e.g., headsets, obtained, e.g., from data.

[00120] The function  $f_{P_b} (\Delta Phase')$  is a smooth function. In one embodiment,

$$[00121] \quad f_{R_b}(\Delta Phase'_b) = \exp \left[ -\frac{\Delta Phase'_b}{Width_{Phase,b}} \right]^2$$

where  $Width_{Phase,b}$  is a width tuning parameter expressed in units of phase. In one embodiment,  $Width_{Phase,b}$  is related to but does not need to be determined from actual data.

5 [00122] For the Coherence probability indicator, no target is used, and in one embodiment,

$$[00123] \quad CPI'_b = \left( \frac{R'_{b21}R'_{b12} + \sigma^2}{R'_{b11}R'_{b22} + \sigma^2} \right)^{CFactor_b}$$

where  $CFactor_b$  is a tuning parameter that may be a constant value in the range of 0.1 to 10; in one embodiment, a value of 0.25 was found to be effective.

10 [00124] FIG. 6 shows one example of the calculation in element 529 of the raw gains, and includes a spatially sensitive voice activity detector (VAD) 621, and a wind activity detector (WAD) 623. Alternate versions of noise reduction may not include the WAD, or the spatially sensitive VAD, and further may not include echo suppression or other reduction. Furthermore, the embodiment shown in FIG. 6 includes additional echo suppression, which may not be included in simpler versions.

15 [00125] In one embodiment, the spatial probability indicators are used to determine what is referred to as the beam gain, a statistical quantity denoted  $BeamGain'_b$  that can be used to estimate the in-beam and out-of-beam power from the total power, e.g., using an out-of-beam spectrum calculator 603, and further, can be used to determine the out-of-beam suppression gain by a spatial suppression gain calculator 611. By convention and in the  
20 embodiments presented herein, the probability indicators are scaled such that the beam gain has a maximum value of 1.

[00126] In one embodiment, the beam gain is

$$[00127] \quad BeamGain'_b = BeamGain_{min} + (1 - BeamGain_{min})RPI'_b \cdot PPI'_b \cdot CPI'_b.$$

[00128] Some embodiments use  $BeamGain_{min}$  of 0.01 to 0.3 (−40dB to −10dB). One  
25 embodiment uses a  $BeamGain_{min}$  of 0.1.

[00129] The in-beam and out-of beam powers are:

[00130]  $Power'_{b,InBeam} = BeamGain'_b{}^2 Y'_b$

[00131]  $Power'_{b,OutOfBeam} = (1 - BeamGain'_b{}^2) Y'_b$ .

[00132] Note that  $Power'_{b,InBeam}$  and  $Power'_{b,OutOfBeam}$  are statistical measures used for  
5 suppression.

[00133] In one version of element 603,

[00134]  $Power'_{b,OutOfBeam} = [0.1 + 0.9(1 - BeamGain'_b{}^2)] Y'_b$ .

[00135] One version of gain calculation uses a spatially-selective noise power spectrum  
calculator 605 that determines an estimate of the noise power (or other metric of the  
10 amplitude) spectrum. One embodiment of the invention uses a leaky minimum follower,  
with a tracking rate determined by at least one leak rate parameter. The leak rate  
parameter need not be the same as for the non-spatially-selective noise estimation used in  
the echo coefficient updating. Denote by  $N'_{b,S}$  the spatially-selective noise spectrum  
estimate. In one embodiment,

15 [00136]  $N'_{b,S} = \min(Power'_{b,OutOfBeam}, (1 + \alpha_b) N'_{b,S_{Prev}})$ ,

where  $N'_{b,S_{Prev}}$  is the already determined, i.e., previous value of  $N'_{b,S}$ . The leak rate  
parameter  $\alpha_b$  is expressed in dB/s such that for a frame time denoted  $T$ ,  $(1 + \alpha_b)^{1/T}$  is  
between 1.2 and 4 if the probability of voice is low, and 1 if the probability of voice is  
high. A nominal value of  $\alpha_b$  is 3dB/s such that  $(1 + \alpha_b)^{1/T} = 1.4$ .

20 [00137] In some embodiments, in order to avoid adding bias to the noise estimate, echo  
gating is used, i.e.,

[00138]  $N'_{b,S} = \min(Power'_{b,OutOfBeam}, (1 + \alpha_b) N'_{b,S_{Prev}})$  if  $N'_{b,S_{Prev}} > 2E'_b$ , else  
 $N'_{b,S} = N'_{b,S_{Prev}}$ .

[00139] That is, the noise estimate is updated only if the previous noise estimate suggests the noise level is greater, e.g., greater than twice the current echo prediction. Otherwise the echo would bias the noise estimate.

[00140] One feature of the noise reducer shown in FIGS. 4, 5 and 6 includes

5 simultaneously suppressing: 1) noise based on a spatially-selective noise estimate, and 2) out-of-beam signals. The gain calculator 529 includes an element 613 to calculates a probability indicator, expressed as a gain for the intermediate signal, e.g., the frequency bins  $Y_n$  based on the spatially-selective estimates of the noise power (or other frequency domain amplitude metric) spectrum, and further on the instantaneous banded input power  
10  $Y'_b$  in a particular band. For simplicity this probability indicator is referred to as a gain, denoted  $Gain_N$ . It should be noted however that this gain  $Gain_N$  is not directly applied, but rather combined with additional gains, i.e., additional probability indicators in a gain combiner 615 to achieve a single gain to apply to achieve a single suppressive action.

[00141] The element 613 is shown with echo suppression, and in some versions does not  
15 include echo suppression.

[00142] An expression found to be effective in terms of computational complexity and effect is given by

[00143] 
$$Gain'_N = \left( \frac{\max(0, Y'_b - \beta'_N N_{b,S})}{Y'_b} \right)^{GainExp}$$

where  $Y'_b$  is the instantaneous banded power (or other frequency domain amplitude  
20 metric),  $N'_{b,S}$  is the banded spatially-selective (out-of-beam) noise estimate, and  $\beta'_N$  is a scaling parameter, typically in the range of 1 to 4. In one version,  $\beta'_N = 1.5$ . The parameter  $GainExp$  is a control of the aggressiveness or rate of transition of the suppression gain from suppression to transmission. This exponent generally takes a value in the range of 0.25 to 4. In one version,  $GainExp = 2$ .

## 25 Adding echo suppression

[00144] Some embodiments of input processing for noise reduction include not only noise suppression, but also simultaneous suppression of echo. In some embodiments of gain calculator 529, element 613 includes echo suppression and in gain calculator 529, the

probability indicator for suppressing echoes is expressed as a gain denoted  $Gain'_{b,N+E}$ . The above noise suppression gain expression, in the case of also including echo suppression, becomes

$$[00145] \quad Gain'_{b,N+E} = \left( \frac{\max(0, Y'_b - \beta'_N N'_{b,S} - \beta'_E E'_b)}{Y'_b} \right)^{GainExp_b} \quad ("Gain I")$$

5 where  $Y'_b$  is again the instantaneous banded power,  $N'_{b,S}, E'_b$  are the banded spatially-selective noise and banded echo estimates, and  $\beta'_N, \beta'_E$  are scaling parameters in the range of 1 to 4, to allow for error in the noise and echo estimates and to offset the gain curve accordingly. Again, they are similar in purpose and magnitude to the constants used in the VAD function, though they are not necessarily the same value. In one embodiment  
10 suitable tuned values are  $\beta'_N = 1.5, \beta'_E = 1.4, GainExp_b = 2$  for all values of  $b$ .

[00146] Several of the expressions for  $Gain'_{N+E}$  described herein have the instantaneous banded input power (or other frequency domain amplitude metric)  $Y'_b$  in both the numerator and denominator. This works well when the banding is properly designed as described herein, with log-like or perceptually spaced frequency bands. In alternate  
15 embodiments of the invention, the denominator uses the estimated banded power spectrum (or other amplitude metric spectrum)  $P'_b$ , so that the above expression for  $Gain'_{b,N+E}$  changes to:

$$[00147] \quad Gain'_{b,N+E} = \left( \frac{\max(0, Y'_b - \beta'_N N'_{b,S} - \beta'_E E'_b)}{P'_b} \right)^{GainExp} \quad ("Gain I_{MOD}").$$

#### Additional independent control of echo suppression

20 [00148] The suppression gain expressions above can be generalized as functions on the domain of the ratio of the instantaneous input power to the expected undesirable signal power, sometimes called "noise" for simplicity. In these gain expressions, the undesirable signal power is the sum of the estimated (location-sensitive) noise power and predicted or estimated echo power. Combining the noise and echo together in this way provides a  
25 single probability indicator in the form of a suppressive gain that causes simultaneous attenuation of both undesirable noise and of undesirable echo.

**[00149]** In some cases, e.g., in cases in which the echo can achieve a level substantially higher than the level of the noise, such suppression may not lead to sufficient echo attenuation. For example, in some applications, there may be a need for only mild reduction of the ambient noise, whilst it is generally required that any echo be suppressed below audibility. To achieve such a desired effect, in one embodiment, an additional scaling of the probability indicator or gain is used, such additional scaling based on the ratio of input audio signal to echo power alone.

**[00150]** Denote by  $f_A(\cdot)$ ,  $f_B(\cdot)$  a pair of suppression gain functions, each having desired properties for suppression gains, e.g., as described above, including, for example being smooth. As one example, each of  $f_A(\cdot)$ ,  $f_B(\cdot)$  has sigmoid function characteristics. In some embodiments, rather than the gain expression being defined as  $f_A\left(\frac{Y'_b}{N'_{b,S} + E'_b}\right)$ , one can instead use a pair of probability indicators, e.g., gains  $f_A\left(\frac{Y'_b}{N'_{b,S}}\right)$ ,  $f_B\left(\frac{Y'_b}{E'_b}\right)$  and determine a combined gain factor from  $f_A\left(\frac{Y'_b}{N'_{b,S}}\right)$  and  $f_B\left(\frac{Y'_b}{E'_b}\right)$ , which allows for independent control of the aggressiveness and depth for the response to noise and echo signal power. In yet another embodiment,  $f_A\left(\frac{Y'_b}{N'_{b,S} + E'_b}\right)$  can be applied for both noise and echo suppression, and  $f_B\left(\frac{Y'_b}{E'_b}\right)$  can be applied for *additional* echo suppression.

**[00151]** In one embodiment the two functions  $f_A\left(\frac{Y'_b}{N'_{b,S}}\right)$ ,  $f_B\left(\frac{Y'_b}{E'_b}\right)$ , or in another embodiment, the two functions  $f_A\left(\frac{Y'_b}{N'_{b,S} + E'_b}\right)$ ,  $f_B\left(\frac{Y'_b}{E'_b}\right)$  are combined as a product to achieve a combined probability indicator, as a suppression gain.

### Combining the suppression gains for simultaneous suppression of out-of-location signals

[00152] In one embodiment, the suppression probability indicator for in-beam signals, expressed as a beam gain 612, called the spatial suppression gain, and denoted  $Gain'_{b,S}$  is  
 5 determined by a spatial suppression gain calculator 611 in element 529 (FIG. 5) as

$$[00153] \quad Gain'_{b,S} = BeamGain'_b = BeamGain_{min} + (1 - BeamGain_{min})RPI'_b \cdot PPI'_b \cdot CPI'_b.$$

[00154] The spatial suppression gain 612 is combined with other suppression gains in gain combiner 615 to form an overall probability indicator expressed as a suppression gain. The overall probability indicator for simultaneous suppression of noise, echo, and out-of-  
 10 beam signals, expressed as a gain  $Gain'_{b,RAW}$ , is in one embodiment the product of the gains:

$$[00155] \quad Gain'_{b,RAW} = Gain'_{b,S} \cdot Gain'_{b,N+E}.$$

[00156] In an alternate embodiment, additional smoothing is applied. In one example embodiment of the gain element 615:

$$15 \quad [00157] \quad Gain'_{b,RAW} = 0.1 + 0.9Gain'_{b,S} \cdot Gain'_{b,N+E}.$$

where the minimum gain 0.1 and  $0.9=(1-0.1)$  factors can be varied for different embodiments to achieve a different minimum value for the gain, with a suggested range of 0.001 to 0.3 (−60dB to −10dB).

[00158] The above expression for  $Gain'_{b,RAW}$  suppresses noise and echo equally. As  
 20 discussed above, it may be desirable to not eliminate noise completely, but to completely eliminate echo. In one such embodiment of gain determination,

$$[00159] \quad Gain'_{b,RAW} = 0.1 + 0.9Gain'_{b,S} \cdot f_A \left( \frac{Y'_b}{N'_{b,S} + E'_b} \right) \cdot f_B \left( \frac{Y'_b}{E'_b} \right),$$

where  $f_A \left( \frac{Y'_b}{N'_{b,S} + E'_b} \right)$  achieves (relatively) modest suppression of both noise and echo,

while  $f_B\left(\frac{Y'_b}{E'_b}\right)$  suppresses the echo more. In a different embodiment,  $f_A(\cdot)$  suppresses only noise, and  $f_B(\cdot)$  suppresses the echo.

[00160] In yet another embodiment,

[00161]  $Gain'_{b,RAW} = 0.1 + 0.9Gain'_{b,S} \cdot Gain'_{b,N+E}$ ,

5 where:

[00162]  $Gain'_{b,E+B} = \left(0.1 + 0.9f_A\left(\frac{Y'_b}{N'_{b,S} + E'_b}\right)\right) \cdot \left(0.1 + 0.9f_B\left(\frac{Y'_b}{E'_b}\right)\right)$ .

[00163] In some embodiments, this noise and echo suppression gain is combined with the spatial feature probability indicator or gain for forming a raw combined gain, and then post-processed by a post-processor 625 and by the post processing step to ensure stability and other desired behavior.

[00164] In another embodiment, the gain function  $f_B\left(\frac{Y'_b}{E'_b}\right)$  specific to the echo suppression is applied as a gain after post-processing by post-processor 625. Some embodiments of gain calculator 529 include a determiner of the additional echo suppression gain and a combiner 627 of the additional echo suppression gain with the post-processed gain to result in the overall  $B$  gains to apply. The inventors discovered that such an embodiment can provide a more specific and deeper attenuation of echo, since the echo probability indicator or gain  $f_B\left(\frac{Y'_b}{E'_b}\right)$  is not subject to the smoothing and continuity imposed by the post-processing.

[00165] FIG. 7 shows a flowchart of a method 700 of operating a processing apparatus 100 to suppress noise and out-of-location signals and in some embodiments echo in a number  $P \geq 1$  of signal inputs 101, e.g., from differently located microphones. In embodiments that include echo suppression, method 700 includes processing a  $Q \geq 1$  reference inputs 102, e.g.,  $Q$  inputs to be rendered on  $Q$  loudspeakers, or signals obtained from  $Q$  loudspeakers.

[00166] In one embodiment, method 700 comprises: accepting 701 in the processing apparatus a plurality of sampled input audio signals 101, and forming 703, 707, 709 a

mixed-down banded instantaneous frequency domain amplitude metric 417 of the input audio signals 101 for a plurality of frequency bands, the forming including transforming 703 into complex-valued frequency domain values for a set of frequency bins. In one embodiment, the forming includes in 703 transforming the input audio signals to frequency bins, downmixing, e.g., beamforming 707 the frequency data, and in 709 banding. In 711, the method includes calculating the power (or other amplitude metric) spectrum of the signal. In alternate embodiments, the downmixing can be before transforming, so that a single mixed-down signal is transformed. In alternate embodiments, the system may make use of an estimate of the banded echo reference, or a similar representation of the frequency domain spectrum of the echo reference provided by another processing component or source within the realized system.

**[00167]** The method includes determining in 705 banded spatial features, e.g., location probability indicators 419 from the plurality of sampled input audio signals.

**[00168]** In embodiments that include simultaneous echo suppression, the method includes accepting 713 one or more reference signals and forming in 715 and 717 a banded frequency domain amplitude metric representation of the one or more reference signals. The representation in one embodiment is the sum. Again in embodiments that include echo suppression, the method includes predicting in 721 a banded frequency domain amplitude metric representation of the echo 415 using adaptively determined echo filter coefficients. The predicting in one embodiment further includes voice-activity detecting—VAD—using the estimate of the banded spectral amplitude metric of the mixed-down signal 413, the estimate of banded spectral amplitude metric of noise, and the previously predicted echo spectral content 415. The coefficients are updated or not according to the results of voice-activity detecting. Updating uses an estimate of the banded spectral amplitude metric of the noise, previously predicted echo spectral content 415, and an estimate of the banded spectral amplitude metric of the mixed-down signal 413. The estimate of the banded spectral amplitude metric of the mixed-down signal is in one embodiment the mixed-down banded instantaneous frequency domain amplitude metric 417 of the input audio signals, while in other embodiments, signal spectral estimation is used.

**[00169]** In some embodiments, the method 700 includes: a) calculating in 723 raw suppression gains including an out-of-location signal gain determined using two or more

of the spatial features 419, and a noise suppression gain determined using spatially-selective noise spectral content; and b) combining the raw suppression gains to a first combined gain for each band. The noise suppression gain in some embodiments includes suppression of echoes, and its calculating 723 also uses the predicted echo spectral content 415.

5 [00170] In some embodiments, the method 700 further includes in 725 carrying out spatially-selective voice activity detection determined using two or more of the spatial features 419 to generate a signal classification, e.g., whether voice or not. In some embodiments, wind detection is used such that the signal classification further includes whether the signal is wind or not.

[00171] The method 700 further includes carrying out post-processing on the first combined gains of the bands to generate a post-processed gain 125 for each band. In some embodiments, the post-processing includes ensuring minimum gain, e.g., in a band dependent manner. One feature of embodiments of the present invention is that the post-processing includes carrying out median filtering of the combined gains, e.g., to ensure there are no outlier gains. Some embodiments of post-processing include ensuring smoothness by carrying out time and/or band-to-band smoothing.

[00172] In some embodiments, the post-processing 725 is according to the signal classification, e.g., whether voice or not, or whether wind or not, and in some embodiments, the characteristics of the median filtering vary according to the signal classification, e.g., whether voice or not, or whether wind or not.

20 [00173] In one embodiment in which echo suppression is included, the method includes calculating in 726 an additional echo suppression gain. In one embodiment, the additional echo suppression gain is included in the first combined gain which is used as a final gain for each band, and in another embodiment, the additional echo suppression gain is combined with the results of applying post-processing to the first combined gain to generate a final gain for each band.

[00174] The method includes applying in 727 the final gain, including interpolating the gain for bin data to carry out suppression on the bin data of the mixed-down signal to form suppressed signal data 133, and applying in 729 one or both of a) output synthesis

and transforming to generate output samples, and b) output remapping to generate output frequency bins.

[00175] Typically,  $P \geq 2$  and  $Q \geq 1$ . However, the methods, systems, and apparatuses disclosed herein can scale down to remain effective for the simpler cases of  $P=1$ ,  $Q \geq 1$  and  $P \geq 2$ ,  $Q=0$ . The methods and apparatuses disclosed herein even work reasonably well for  $P=1$ ,  $Q=0$ . Although this final example is a reduced and perhaps trivial embodiment of the presented invention, it is noted that the ability of the proposed framework to scale is advantageous, and furthermore the lower signal operation case may be required in practice should one or more of the input audio signals or reference signals become corrupted or unavailable, e.g. due to the failure of a sensor or microphone.

[00176] Whilst the disclosure is presented for a complete noise reduction method (FIG. 7), system or apparatus (FIGS. 5, 6,) that includes all aspects of suppression, including simultaneous echo, noise, and out-of-spatial-location suppression, or presented as a computer-readable storage medium that includes instructions that when executed by one or more processors of a processing system (see FIG. 8 described below), cause a processing apparatus that includes the processing system to carry out the method such as that of FIG. 7, note that the example embodiments also provide a scalable solution for simpler applications and situations. Furthermore, noise reduction is only one example of input processing that determines gains that can be post-processed by the post-processing method that includes median filtering described in embodiments of the present invention.

### *A processing system-based apparatus*

[00177] FIG. 8 shows a simplified block diagram of one processing apparatus embodiment 800 for processing one or more of audio inputs 101, e.g., from microphones (not shown). The processing apparatus 800 is to determine a set of gains, to post-process the gains including median filtering the determined gains, and to generate audio output 137 that has been modified by application of the gains. One version achieves one or more of perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization that takes into account the variation in the perception of audio depending on the reproduction level of the audio signal. Another version achieved noise reduction.

[00178] One noise reduction version includes echo reduction, and in such a version, the processing apparatus also accepts one or more reference signals 103, e.g., from one or more loudspeakers (not shown) or from the feed(s) to such loudspeaker(s). In one such noise reduction version, the processing apparatus 800 is to generate audio output 137 that  
5 has been modified by suppressing, in one embodiment noise and out-of-location signals, and in another embodiment also echoes as specified in accordance to one or more features of the present invention. The apparatus, for example, can implement the system shown in FIG. 6, and any alternates thereof, and can carry out, when operating, the method of FIG. 7 including any variations of the method described herein. Such an apparatus may be  
10 included, for example, in a headphone set such as a Bluetooth headset. The audio inputs 101, the reference input(s) 103 and the audio output 137 are assumed to be in the form of frames of  $M$  samples of sampled data. In the case of analog input, a digitizer including an analog-to-digital converter and quantizer would be present. For audio playback, a de-quantizer and a digital-to-analog converter would be present. Such and other elements that  
15 might be included in a complete audio processing system, e.g., a headset device are left out, and how to include such elements would be clear to one skilled in the art.

[00179] The embodiment shown in FIG. 8 includes a processing system 803 that is configured in operation to carry out the suppression methods described herein. The processing system 803 includes at least one processor 805, which can be the processing  
20 unit(s) of a digital signal processing device, or a CPU of a more general purpose processing device. The processing system 803 also includes a storage subsystem 807 typically including one or more memory elements. The elements of the processing system are coupled, e.g., by a bus subsystem or some other interconnection mechanism not shown in FIG. 8. Some of the elements of processing system 803 may be integrated into a  
25 single circuit, using techniques commonly known to one skilled in the art.

[00180] The storage subsystem 807 includes instructions 811 that when executed by the processor(s) 805, cause carrying out of the methods described herein.

[00181] In some embodiments, the storage subsystem 807 is configured to store one or more tuning parameters 813 that can be used to vary some of the processing steps carried  
30 out by the processing system 803.

[00182] The system shown in FIG. 8 can be incorporated in a specialized device such as a headset, e.g., a wireless Bluetooth headset. The system also can be part of a general purpose computer, e.g., a personal computer configured to process audio signals.

#### Voice activity detection with settable sensitivity

5 [00183] In some embodiments of the invention, the post-processing, e.g., the median filtering is controlled by signal classification as determined by a VAD. The invention is not limited to any particular type of VAD, and many VADs are known in the art. When applied to suppression, the inventors have discovered that suppression works best when different parts of the suppression system are controlled by different VADs, each such

10 VAD custom designed for the functions of the suppressor in which it is used in, rather than having an “optimal” VAD for all uses. Therefore, in some versions of the input processing for noise reduction, a plurality of VADs, each controlled by a small set of tuning parameters that separately control sensitivity and selectivity, including spatial selectivity, such parameters tuned according to the suppression elements in which the

15 VAD is used. Each of the plurality of the VADs is an instantiation of a universal VAD that determines indications of voice activity from  $Y'_b$ . The universal VAD is controlled by a set of parameters and uses an estimate of noise spectral content, the banded frequency domain amplitude metric representation of the echo, and the banded spatial features. The set of parameters includes whether the estimate of noise spectral content is spatially

20 selective or not. The type of indication of voice activity that a particular instantiation determines is controlled by a selection of the parameters.

[00184] One embodiment of a general spatially-selective VAD structure—the universal VAD to calculate voice activity that can be tuned for various functions—is

$$[00185] \quad S = \sum_{b=1}^B (BeamGain'_b)^{BeamGainExp} \left( \frac{\max(0, Y'_b - \beta_{bN} \cdot (N'_b \vee N'_{b,S}) - \beta_{bE} E'_b)}{Y'_b + Y'_{b,sens}} \right),$$

25 where  $BeamGain'_b = BeamGain_{min} + (1 - BeamGain_{min}) RPI'_b \cdot PPI'_b \cdot CPI'_b$ ,  
 $BeamGainExp$  is a parameter that for larger values increases the aggressiveness of the spatial selectivity of the VAD, and is 0 for a non-spatially-selective VAD,  $N'_b \vee N'_{b,S}$  denotes either the total noise power (or other frequency domain amplitude metric) estimate  $N'_b$ , or the spatially-selective noise estimate  $N'_{b,S}$  determined using the out-of-

beam power (or other frequency domain amplitude metric),  $\beta_N, \beta_E > 1$  are margins for noise end echo, respectively and  $Y'_{sens}$  is a settable sensitivity offset. The values of  $\beta_N, \beta_E$  are between 1 and 4. *BeamGainExp* is between 0.5 to 2.0 when spatial selectivity is desired, and is 1.5 for one embodiment of a spatially-selective VAD, e.g., used to control post-processing in some embodiments of the invention.  $RPI'_b, PPI'_b$ , and  $CPI'_b$  are, as above, three spatial probability indicators, namely the ratio probability indicator, the phase probability indicator, and the coherence probability indicator.

**[00186]** The above expression also controls the operation of the universal voice activity detecting method.

**[00187]** For any given set of parameters to generate the voice indicator value  $S$  a binary decision or classifier can be obtained by considering the test  $S > S_{thresh}$  as indicating the presence of voice. It should also be apparent that the value  $S$  can be used as a continuous indicator of the instantaneous voice level. Furthermore, an improved useful universal VAD for operations such as transmission control or controlling the post processing could be obtained using a suitable “hang over” or period of continued indication of voice after a detected event. Such a hang over period may vary from 0 to 500ms, and in one embodiment a value of 200ms was used. During the hang over period, it can be useful to reduce the activation threshold, for example by a factor of 2/3. This creates increased sensitivity to voice and stability once a talk burst has commenced.

**[00188]** For spatially-selective voice activity detection to control one or more post-processing operations, e.g., for a spatially-selective VAD, the noise in the above expression is  $N'_{b,S}$  determined using an out-of-beam estimate of power (or other frequency domain amplitude metric).  $Y_{sens}$  is set to be around expected microphone and system noise level, obtained by experiments on typical components.

## 25 *General*

**[00189]** Unless specifically stated otherwise, it is appreciated that throughout the specification discussions using terms such as “generating,” “processing,” “computing,” “calculating,” “determining” or the like, may refer to, without limitation, the action and/or processes of hardware, e.g., an electronic circuit, a computer or computing system, or

similar electronic computing device, that manipulate and/or transform data represented as physical, such as electronic, quantities into other data similarly represented as physical quantities.

**[00190]** In a similar manner, the term “processor” may refer to any device or portion of a device that processes electronic data, e.g., from registers and/or memory to transform that electronic data into other electronic data that, e.g., may be stored in registers and/or memory. A “computer” or a “computing machine” or a “computing platform” may include one or more processors.

**[00191]** Note that when a method is described that includes several elements, e.g., several steps, no ordering of such elements, e.g., of such steps is implied, unless specifically stated.

**[00192]** The methodologies described herein are, in some embodiments, performable by one or more processors that accept logic, instructions encoded on one or more computer-readable media. When executed by one or more of the processors, the instructions cause carrying out at least one of the methods described herein. Any processor capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken is included. Thus, one example is a typical processing system that includes one or more processors. Each processor may include one or more of a CPU or similar element, a graphics processing unit (GPU), field-programmable gate array, application-specific integrated circuit, and/or a programmable DSP unit. The processing system further includes a storage subsystem with at least one storage medium, which may include memory embedded in a semiconductor device, or a separate memory subsystem including main RAM and/or a static RAM, and/or ROM, and also cache memory. The storage subsystem may further include one or more other storage devices, such as magnetic and/or optical and/or further solid state storage devices. A bus subsystem may be included for communicating between the components. The processing system further may be a distributed processing system with processors coupled by a network, e.g., via network interface devices or wireless network interface devices. If the processing system requires a display, such a display may be included, e.g., a liquid crystal display (LCD), organic light emitting display (OLED), or a cathode ray tube (CRT) display. If manual data entry is required, the processing system also includes an input device such as one or more of an alphanumeric input unit such as a keyboard, a pointing control device such as a mouse,

and so forth. The term storage device, storage subsystem, or memory unit as used herein, if clear from the context and unless explicitly stated otherwise, also encompasses a storage system such as a disk drive unit. The processing system in some configurations may include a sound output device, and a network interface device.

5   **[00193]**   In some embodiments, a non-transitory computer-readable medium is configured with, e.g., encoded with instructions, e.g., logic that when executed by one or more processors of a processing system such as a digital signal processing device or subsystem that includes at least one processor element and a storage subsystem, cause carrying out a method as described herein. Some embodiments are in the form of the logic itself. A non-  
10   transitory computer-readable medium is any computer-readable medium that is not specifically a transitory propagated signal or a transitory carrier wave or some other transitory transmission medium. The term “non-transitory computer-readable medium” thus covers any tangible computer-readable storage medium. Non-transitory computer-readable media include any tangible computer-readable storage media and may take many  
15   forms including non-volatile storage media and volatile storage media. Non-volatile storage media include, for example, static RAM, optical disks, magnetic disks, and magneto-optical disks. Volatile storage media includes dynamic memory, such as main memory in a processing system, and hardware registers in a processing system. In a typical processing system as described above, the storage subsystem thus a computer-  
20   readable storage medium that is configured with, e.g., encoded with instructions, e.g., logic, e.g., software that when executed by one or more processors, causes carrying out one or more of the method steps described herein. The software may reside in the hard disk, or may also reside, completely or at least partially, within the memory, e.g., RAM and/or within the processor registers during execution thereof by the computer system.  
25   Thus, the memory and the processor registers also constitute a non-transitory computer-readable medium on which can be encoded instructions to cause, when executed, carrying out method steps.

**[00194]**   While the computer-readable medium is shown in an example embodiment to be a single medium, the term "medium" should be taken to include a single medium or  
30   multiple media (e.g., several memories, a centralized or distributed database, and/or associated caches and servers) that store the one or more sets of instructions.

**[00195]** Furthermore, a non-transitory computer-readable medium, e.g., a computer-readable storage medium may form a computer program product, or be included in a computer program product.

**[00196]** In alternative embodiments, the one or more processors operate as a standalone device or may be connected, e.g., networked to other processor(s), in a networked deployment, or the one or more processors may operate in the capacity of a server or a client machine in server-client network environment, or as a peer machine in a peer-to-peer or distributed network environment. The term processing system encompasses all such possibilities, unless explicitly excluded herein. The one or more processors may form a personal computer (PC), a media playback device, a headset device, a hands-free communication device, a tablet PC, a set-top box (STB), a personal digital assistant (PDA), a game machine, a cellular telephone, a Web appliance, a network router, switch or bridge, or any machine capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken by that machine.

**[00197]** Note that while some diagram(s) only show(s) a single processor and a single storage subsystem, e.g., a single memory that stores the logic including instructions, those skilled in the art will understand that many of the components described above are included, but not explicitly shown or described in order not to obscure the inventive aspect. For example, while only a single machine is illustrated, the term “machine” shall also be taken to include any collection of machines that individually or jointly execute a set (or multiple sets) of instructions to perform any one or more of the methodologies discussed herein.

**[00198]** Thus, as will be appreciated by those skilled in the art, embodiments of the present invention may be embodied as a method, an apparatus such as a special purpose apparatus, an apparatus such as a data processing system, logic, e.g., embodied in a non-transitory computer-readable medium, or a computer-readable medium that is encoded with instructions, e.g., a computer-readable storage medium configured as a computer program product. The computer-readable medium is configured with a set of instructions that when executed by one or more processors cause carrying out method steps. Accordingly, aspects of the present invention may take the form of a method, an entirely hardware embodiment, an entirely software embodiment or an embodiment combining software and hardware aspects. Furthermore, the present invention may take the form of

program logic, e.g., a computer program on a computer-readable storage medium, or the computer-readable storage medium configured with computer-readable program code, e.g., a computer program product.

**[00199]** It will also be understood that embodiments of the present invention are not  
5 limited to any particular implementation or programming technique and that the invention may be implemented using any appropriate techniques for implementing the functionality described herein. Furthermore, embodiments are not limited to any particular programming language or operating system.

**[00200]** Reference throughout this specification to “one embodiment” or “an embodiment”  
10 means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases “in one embodiment” or “in an embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined  
15 in any suitable manner, as would be apparent to one of ordinary skill in the art from this disclosure, in one or more embodiments.

**[00201]** Similarly it should be appreciated that in the above description of example  
embodiments of the invention, various features of the invention are sometimes grouped  
together in a single embodiment, figure, or description thereof for the purpose of  
20 streamlining the disclosure and aiding in the understanding of one or more of the various inventive aspects. This method of disclosure, however, is not to be interpreted as reflecting an intention that the claimed invention requires more features than are expressly recited in each claim. Rather, as the following claims reflect, inventive aspects lie in less than all features of a single foregoing disclosed embodiment. Thus, the claims following  
25 the DESCRIPTION OF EXAMPLE EMBODIMENTS are hereby expressly incorporated into this DESCRIPTION OF EXAMPLE EMBODIMENTS, with each claim standing on its own as a separate embodiment of this invention.

**[00202]** Furthermore, while some embodiments described herein include some but not  
other features included in other embodiments, combinations of features of different  
30 embodiments are meant to be within the scope of the invention, and form different

embodiments, as would be understood by those skilled in the art. For example, in the following claims, any of the claimed embodiments can be used in any combination.

**[00203]** Furthermore, some of the embodiments are described herein as a method or combination of elements of a method that can be implemented by a processor of a computer system or by other means of carrying out the function. Thus, a processor with the necessary instructions for carrying out such a method or element of a method forms a means for carrying out the method or element of a method. Furthermore, an element described herein of an apparatus embodiment is an example of a means for carrying out the function performed by the element for the purpose of carrying out the invention.

**[00204]** In the description provided herein, numerous specific details are set forth. However, it is understood that embodiments of the invention may be practiced without these specific details. In other instances, well-known methods, structures and techniques have not been shown in detail in order not to obscure an understanding of this description.

**[00205]** As used herein, unless otherwise specified, the use of the ordinal adjectives “first”, “second”, “third”, etc., to describe a common object, merely indicate that different instances of like objects are being referred to, and are not intended to imply that the objects so described must be in a given sequence, either temporally, spatially, in ranking, or in any other manner.

**[00206]** While in one embodiment, the short time Fourier transform (STFT) is used to obtain the frequency bands, the invention is not limited to the STFT. Transforms such as the STFT are often referred to as circulant transforms. Most general forms of circulant transforms can be represented by buffering, a window, a twist (real value to complex value transformation) and a DFT, e.g., FFT. A complex twist after the DFT can be used to adjust the frequency domain representation to match specific transform definitions. The invention may be implemented by any of this class of transforms, including the modified DFT (MDFT), the short time Fourier transform (STFT), and with a longer window and wrapping, a conjugate quadrature mirror filter (CQMF). Other standard transforms such as the Modified discrete cosine transform (MDCT) and modified discrete sine transform (MDST), can also be used, with an additional complex twist of the frequency domain bins, which does not change the underlying frequency resolution or processing ability of

the transform and thus can be left until the end of the processing chain, and applied in the remapping if required.

**[00207]** All U.S. patents, U.S. patent applications, and International (PCT) patent applications designating the United States cited herein are hereby incorporated by reference, except in those jurisdictions that do not permit incorporation by reference, in which case the Applicant reserves the right to insert any portion of or all such material into the specification by amendment without such insertion considered new matter. In the case the Patent Rules or Statutes do not permit incorporation by reference of material that itself incorporates information by reference, the incorporation by reference of the material herein excludes any information incorporated by reference in such incorporated by reference material, unless such information is explicitly incorporated herein by reference.

**[00208]** Any discussion of other art in this specification should in no way be considered an admission that such art is widely known, is publicly known, or forms part of the general knowledge in the field at the time of invention.

**[00209]** In the claims below and the description herein, any one of the terms comprising, comprised of or which comprises is an open term that means including at least the elements/features that follow, but not excluding others. Thus, the term comprising, when used in the claims, should not be interpreted as being limitative to the means or elements or steps listed thereafter. For example, the scope of the expression a device comprising A and B should not be limited to devices consisting of only elements A and B. Any one of the terms including or which includes or that includes as used herein is also an open term that also means including at least the elements/features that follow the term, but not excluding others. Thus, including is synonymous with and means comprising.

**[00210]** Similarly, it is to be noticed that the term coupled, when used in the claims, should not be interpreted as being limitative to direct connections only. The terms “coupled” and “connected,” along with their derivatives, may be used. It should be understood that these terms are not intended as synonyms for each other, but may be. Thus, the scope of the expression “a device A coupled to a device B” should not be limited to devices or systems wherein an input or output of device A is directly connected to an output or input of device B. It means that there exists a path between device A and device B which may be a path including other devices or means in between. Furthermore, “coupled to” does not

imply direction. Hence, the expression “a device A is coupled to a device B” may be synonymous with the expression “a device B is coupled to a device A.” “Coupled” may mean that two or more elements are either in direct physical or electrical contact, or that two or more elements are not in direct contact with each other but yet still co-operate or  
5 interact with each other.

**[00211]** In addition, use of the “a” or “an” are used to describe elements and components of the embodiments herein. This is done merely for convenience and to give a general sense of the invention. This description should be read to include one or at least one and the singular also includes the plural unless it is obvious that it is meant otherwise.

10 **[00212]** Thus, while there has been described what are believed to be the preferred embodiments of the invention, those skilled in the art will recognize that other and further modifications may be made thereto without departing from the spirit of the invention, and it is intended to claim all such changes and modifications as fall within the scope of the invention. For example, any formulas given above are merely representative of procedures  
15 that may be used. Functionality may be added or deleted from the block diagrams and operations may be interchanged among functional blocks. Steps may be added to or deleted from methods described within the scope of the present invention.

## CLAIMS

We claim:

1. A method of applying post-processing to raw banded gains to improve the raw banded gains for applying to one or more audio signal to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization, the method comprising:

applying post-processing to raw banded gains to generate banded post-processed gains to apply to the one or more audio signals, the raw banded gains determined by input processing one or more input audio signals to generate the raw banded gains at a plurality of frequency bands comprising one or more frequency bins, the raw banded gains being in order to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization,

wherein the generating of a particular post-processed gain for a particular frequency band includes at least median filtering using raw gain values for frequency bands adjacent to the particular frequency band,

wherein the post-processing is according to one or more properties, including an end condition and a width for the median filtering, and

wherein at least one property of the post-processing depends on signal classification of the one or more input audio signals.

2. A method of processing one or more input audio signals, the method comprising:

input processing one or more input audio signals to generate raw banded gains at a plurality of frequency bands comprising one or more frequency bins, the raw banded gains being in order to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization; and

applying post-processing to the raw banded gains to generate banded post-processed gains to apply to the one or more input audio signals,

wherein the generating of a particular post-processed gain for a particular frequency band includes at least median filtering using raw gain values for frequency bands adjacent to the particular frequency band,

wherein the post-processing is according to one or more properties, including an end condition and a width for the median filtering, and

wherein at least one property of the post-processing depends on signal classification of the one or more input audio signals.

- 10 3. A method as recited in claims 1 or 2, wherein the post-processing further comprises at least one of frequency-band-to-frequency-band smoothing and smoothing across time.
4. A method as recited in claim 3, wherein at least one of the frequency-band-to-frequency-band smoothing and the smoothing across time depends upon signal classification.
- 15 5. A method as recited in any one of claims 1 to 4, wherein the signal classification includes whether the one or more input audio signals are likely or not to be wind.
6. A method as recited in any one of claims 1 to 4, wherein at least one of the end condition and the width of the median filtering depends on the signal classification.
7. A method as recited in claim 6, wherein the signal classification includes whether the one or more input audio signals are likely or not to be voice.
- 20 8. A method as recited in claim 6, wherein the signal classification includes whether the one or more input audio signals are likely or not to be noise.
9. A method as recited in any one of claims 1 to 8, wherein the frequency bands are on a perceptual or logarithmic scale.
- 25 10. A method as recited in any one of claims 1 to 9, wherein the input processing is to determine the raw banded gains for reducing noise.

11. A method as recited in any one of claims 1 to 9, wherein the input processing is to determine the raw banded gains from more than one input audio signal for reducing noise and out-of-location signals.
12. A method as recited in any one of claims 1 to 9, wherein the input processing is to  
5 determine the raw banded gains from one or more input audio signals and one or more reference signals, the determined gains being for reducing noise and echoes.
13. A method as recited in any one of claims 1 to 12, wherein the input processing is to  
10 determine the raw banded gains for one or more of perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization.
14. A tangible computer-readable storage medium comprising instructions that when executed by at least one processor of a processing system, cause carrying out a method as recited in any one of claims 1 to 13.
15. Program logic embodied in a tangible computer-readable storage medium, wherein the  
15 logic when executed by at least one processor causes carrying out a method as recited in any one of claims 1 to 13.
16. An apparatus to post-process raw banded gains for applying to an audio signal, the raw banded gains determined by input processing one or more input audio signals, the apparatus comprising one or more processors and a tangible computer-readable  
20 storage medium comprising instructions that when executed by at least one processor of the one or more processors, cause carrying out a method as recited in any one of claims 1 to 13.
17. An apparatus to post-process raw banded gains for applying to an audio signal, the apparatus comprising:  
25 a post-processor accepting raw banded gains determined by input processing one or more input audio signals by an input processor, the post-processor applying post-processing to the raw banded gains to generate banded post-processed gains to apply to the one or more input audio signals, the input processing to generate the raw banded gains at a plurality of frequency bands comprising one or more frequency bins, the  
30 raw banded gains being in order to carry out one or more of reducing noise, reducing

out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization,

wherein the banded post-processed gains are for applying to the one or more input audio signals,

wherein the post-processor includes a median filter to carry out median filtering of the raw banded gains,

wherein the generating by the post-processor of a particular post-processed gain for a particular frequency band includes the median filtering using raw gain values for frequency bands adjacent to the particular frequency band,

wherein the post-processing is according to one or more properties, including an end condition and a width for the median filtering, and

wherein at least one property of the post-processing depends on signal classification of the one or more input audio signals.

18. An apparatus to process one or more input audio signals; the apparatus comprising:

an input processor accepting the one or more input audio signals and input processing the one or more input audio signals to generate raw banded gains at a plurality of frequency bands comprising one or more frequency bins, the raw banded gains for carrying out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization;

a post-processor accepting the raw banded gains and applying post-processing to the raw banded gains to generate post-processed raw banded gains to apply to the one or more input signals,

wherein the post-processor includes a median filter to carry out median filtering of the raw banded gains,

wherein the generating by the post-processor of a particular post-processed gain for a particular frequency band includes the median filtering using raw gain values for frequency bands adjacent to the particular frequency band,

wherein the post-processing is according to one or more properties, including an end condition and a width for the median filtering, and

wherein at least one property of the post-processing depends on signal classification of the one or more input audio signals.

- 5 19. An apparatus as recited in claims 17 or 18, wherein the post-processor includes a smoothing filter to smooth the median filtered gains, including at least one of frequency-band-to-frequency-band smoothing and smoothing across time.
20. An apparatus as recited in any one of claims 17 to 19, further comprising a signal classifier to generate a signal classification of the one or more input audio signals,  
10 wherein the width of the median filtering depends on the signal classification of the one or more input audio signals.
21. An apparatus as recited in claim 20, wherein the signal classifier includes a voice activity detector such that the signal classification includes whether the input audio signals are likely or not to be voice.
- 15 22. An apparatus as recited in any one of claims 17 to 21, wherein the width of the median filtering depends on the spectral flux of the one or more input audio signals.
23. An apparatus as recited in any one of claims 17 to 22, wherein the width of the median filtering for the particular frequency band depends on the particular frequency band.
24. An apparatus as recited in any one of claims 17 to 23, wherein the frequency bands are  
20 on a perceptual or logarithmic scale.
25. An apparatus as recited in any one of claims 17 to 24, wherein the median filtering depends on one or more of a classification of the one or more input audio signals.
26. An apparatus as recited in any one of claims 17 to 25, wherein the raw banded gains determined by the input processing are for reducing noise.
- 25 27. An apparatus as recited in any one of claims 17 to 25, wherein the raw banded gains determined by the input processing are determined from more than one input audio signal and are for reducing noise and out-of-location signals.

28. An apparatus as recited in any one of claims 17 to 25, wherein the input processor further accepts one or more reference signals, and wherein the raw banded gains determined by the input processing are determined from one or more input audio signals and the one or more reference signals, and are for reducing noise and echoes.

5 29. An apparatus as recited in any one of claims 17 to 28, wherein the raw banded gains determined by the input processing are for one or more of perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization.

10 30. A system for post-processing raw banded gains to generate banded post-processed gains for applying to an audio signal, the system comprising:

means for post-processing raw banded gains to generate banded post-processed gains, the raw banded gains determined by a means for input processing one or more input audio signals to generate the raw banded gains at a plurality of frequency bands comprising one or more frequency bins, the raw banded gains being in order to carry  
15 out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization,

wherein the banded post-processed gains are applied to the one or more audio signals  
20 to carry out the one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization,

wherein the generating of a particular post-processed gain for a particular frequency band includes at least median filtering using raw gain values for frequency bands adjacent to the particular frequency band,

25 wherein the post-processing applied by the means for post-processing is according to one or more properties, including an end condition and a width for the median filtering, and

wherein at least one property of the post-processing depends on signal classification of the one or more input audio signals.

30 31. A system for processing one or more input audio signals, the system comprising:

means for input processing one or more input audio signals to generate raw banded gains at a plurality of frequency bands comprising one or more frequency bins, the raw banded gains being in order to carry out one or more of reducing noise, reducing out-of-location signals, reducing echoes, perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization;

means for post-processing raw banded gains to generate banded post-processed gains to apply to the one or more input audio signals,

wherein the generating of a particular post-processed gain for a particular frequency band includes at least median filtering using raw gain values for frequency bands adjacent to the particular frequency band,

wherein the post-processing applied by the means for post-processing is according to one or more properties, including an end condition and a width for the median filtering, and

wherein at least one property of the post-processing depends on signal classification of the one or more input audio signals.

32. A system as recited in claim 30 or 31, wherein the post-processing further comprises at least one of frequency-band-to-frequency-band smoothing and smoothing across time.

33. A system as recited in claim 32, wherein at least one of the frequency-band-to-frequency-band smoothing and the smoothing across time depends upon signal classification.

34. A system as recited in any one of claims 30 to 33, wherein the signal classification includes whether the one or more input audio signals are likely or not to be wind.

35. A system as recited in any one of claims 30 to 33, wherein at least one of the end condition and the width of the median filtering depends on the signal classification.

36. A system as recited in claim 35, wherein the signal classification includes whether the one or more input audio signals are likely or not to be voice.

37. A system as recited in claim 35, wherein the signal classification includes whether the one or more input audio signals are likely or not to be noise.
38. A system as recited in any one of claims 30 to 37, wherein the frequency bands are on a perceptual or logarithmic scale.
- 5 39. A system as recited in any one of claims 30 to 38, wherein the input processing is to determine the raw banded gains for reducing noise.
40. A system as recited in any one of claims 30 to 38, wherein the input processing is to determine the raw banded gains from more than one input audio signal for reducing noise and out-of-location signals.
- 10 41. A system as recited in any one of claims 30 to 38, wherein the input processing is to determine the raw banded gains from one or more input audio signals and one or more reference signals, the determined gains being for reducing noise and echoes.
42. A system as recited in any one of claims 30 to 41, wherein the input processing is to determine the raw banded gains for one or more of perceptual domain-based leveling, perceptual domain-based dynamic range control, and perceptual domain-based dynamic equalization.
- 15

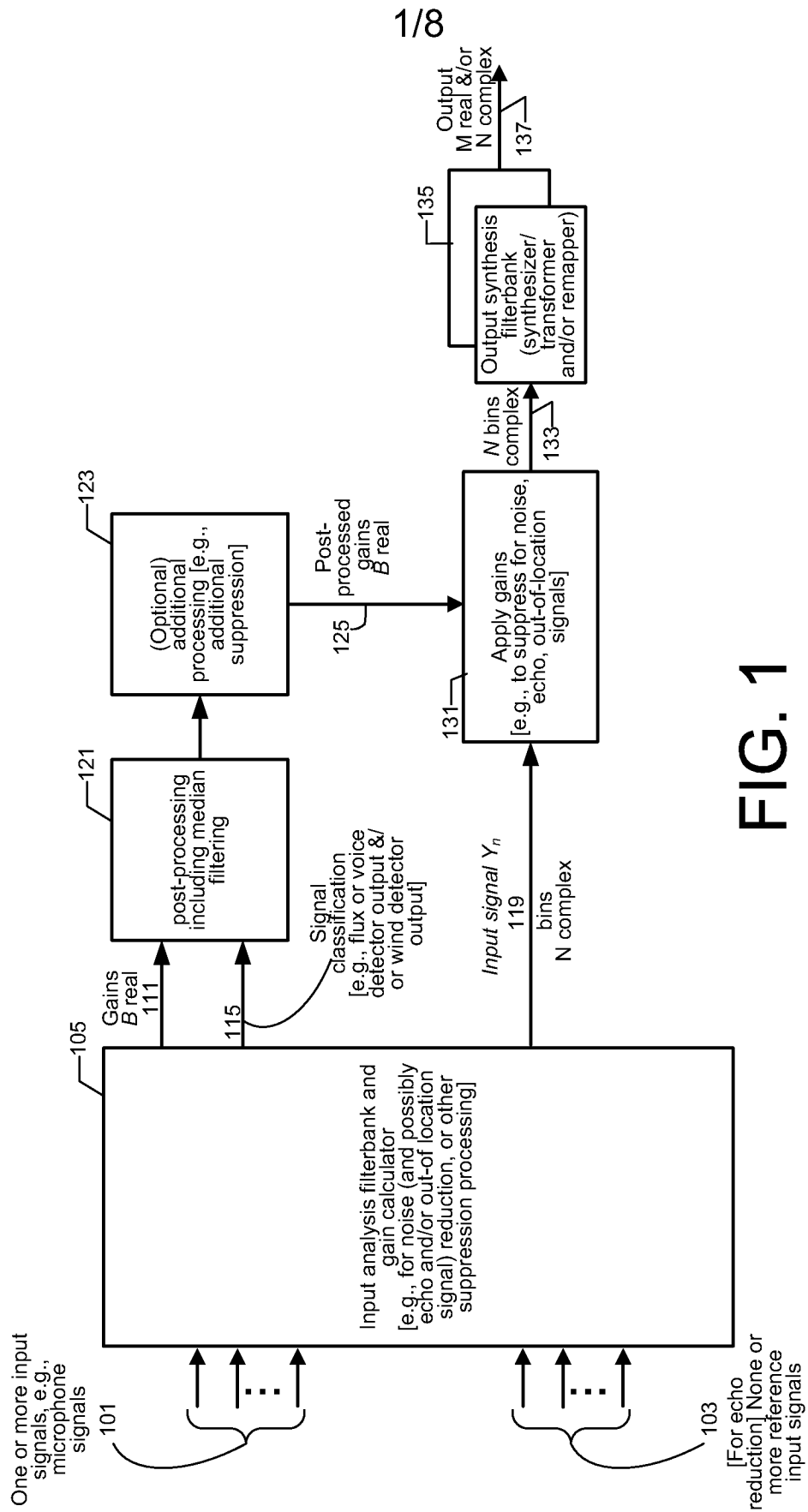


FIG. 1

2/8

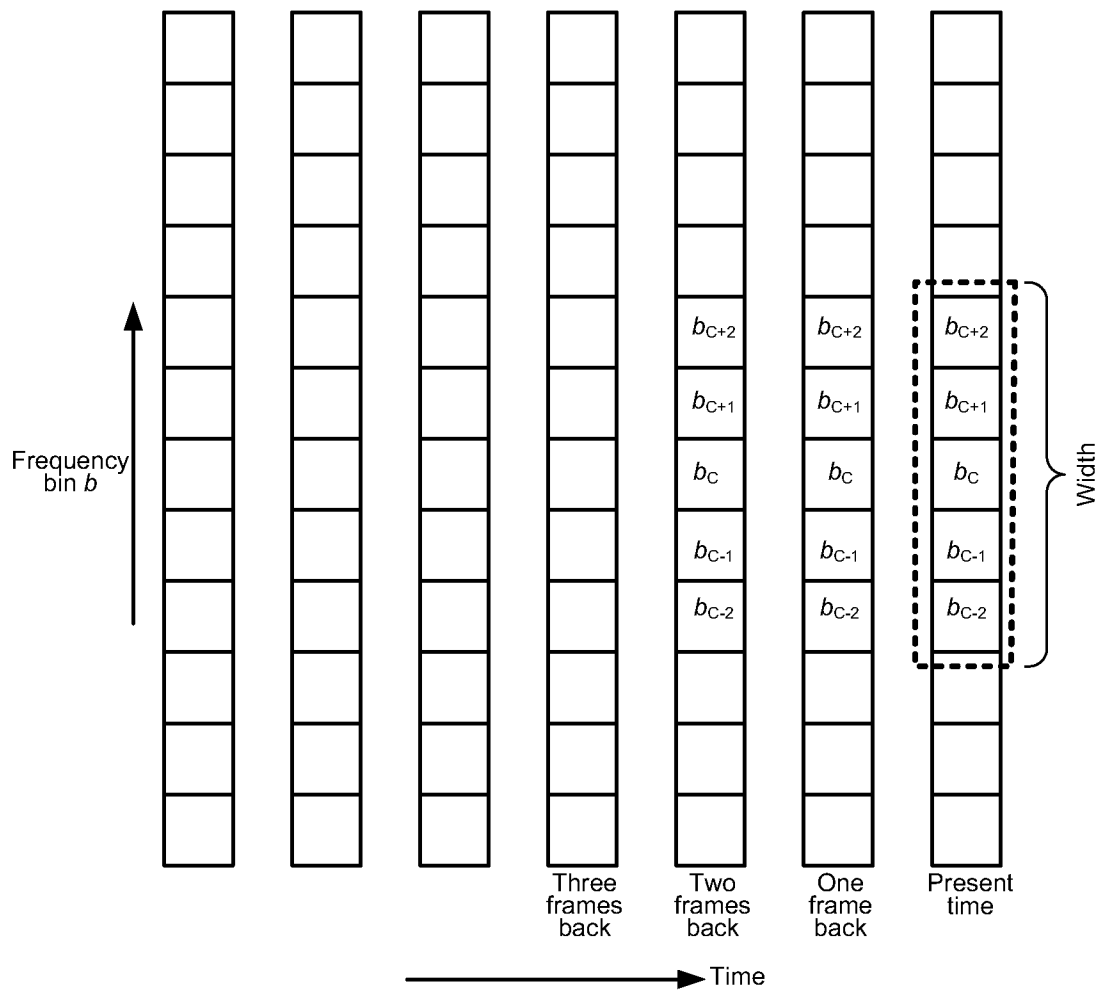


FIG. 2

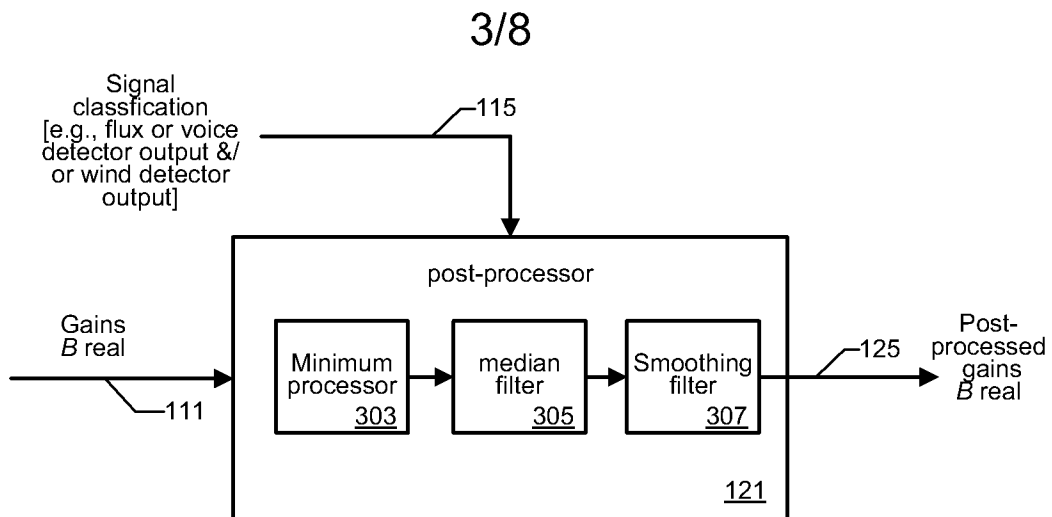


FIG. 3A

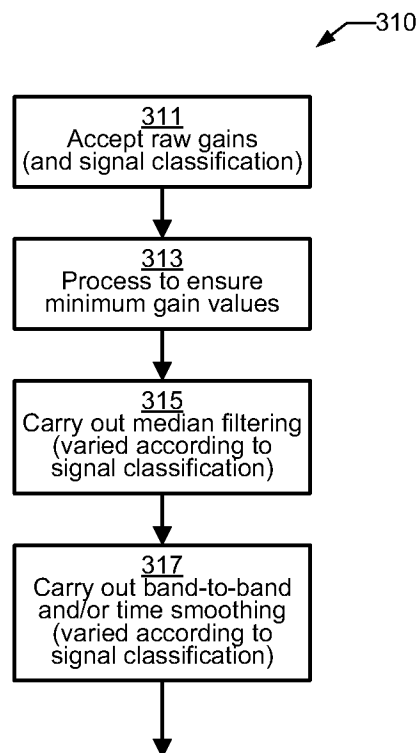


FIG. 3B

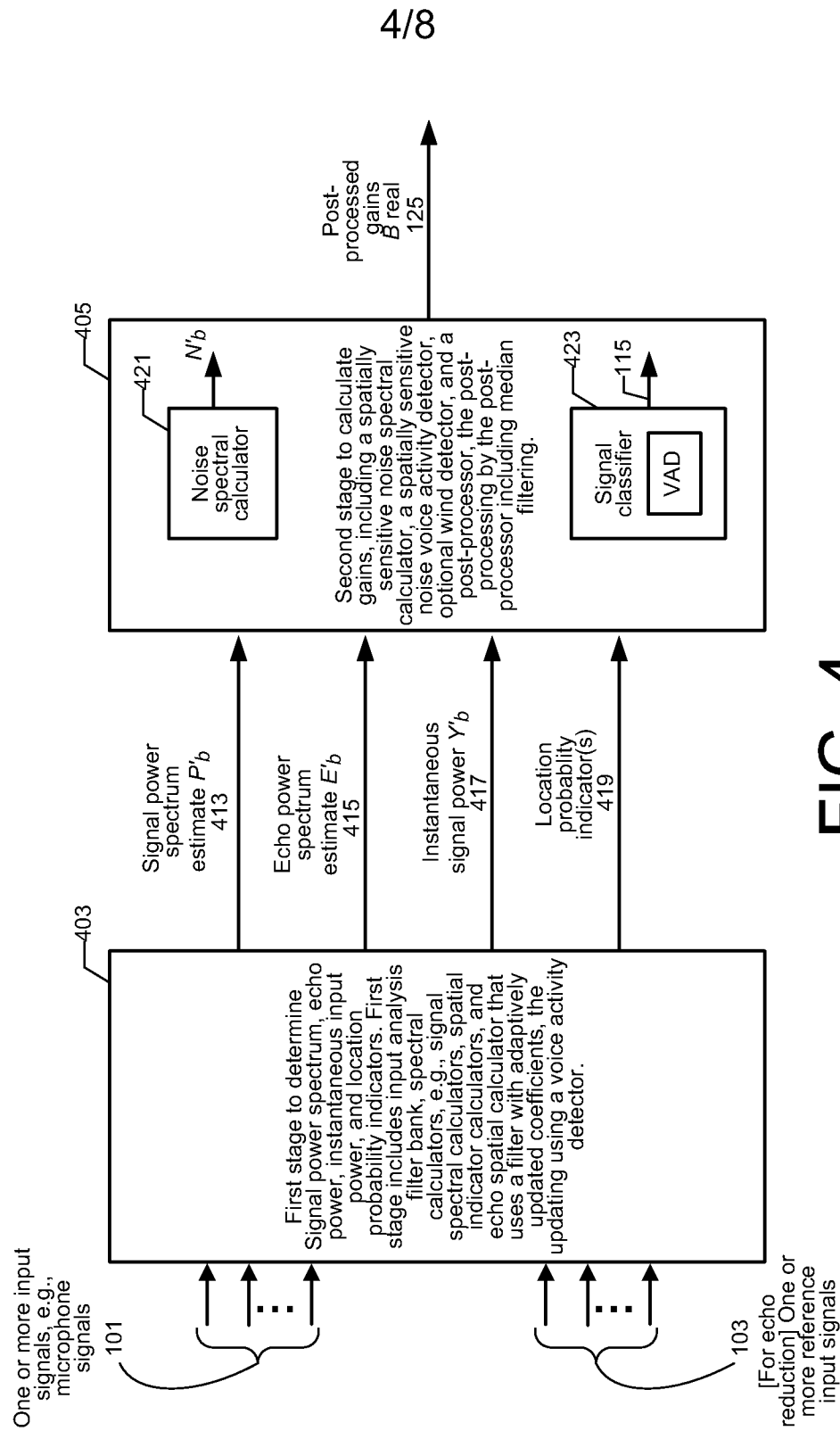


FIG. 4

5/8

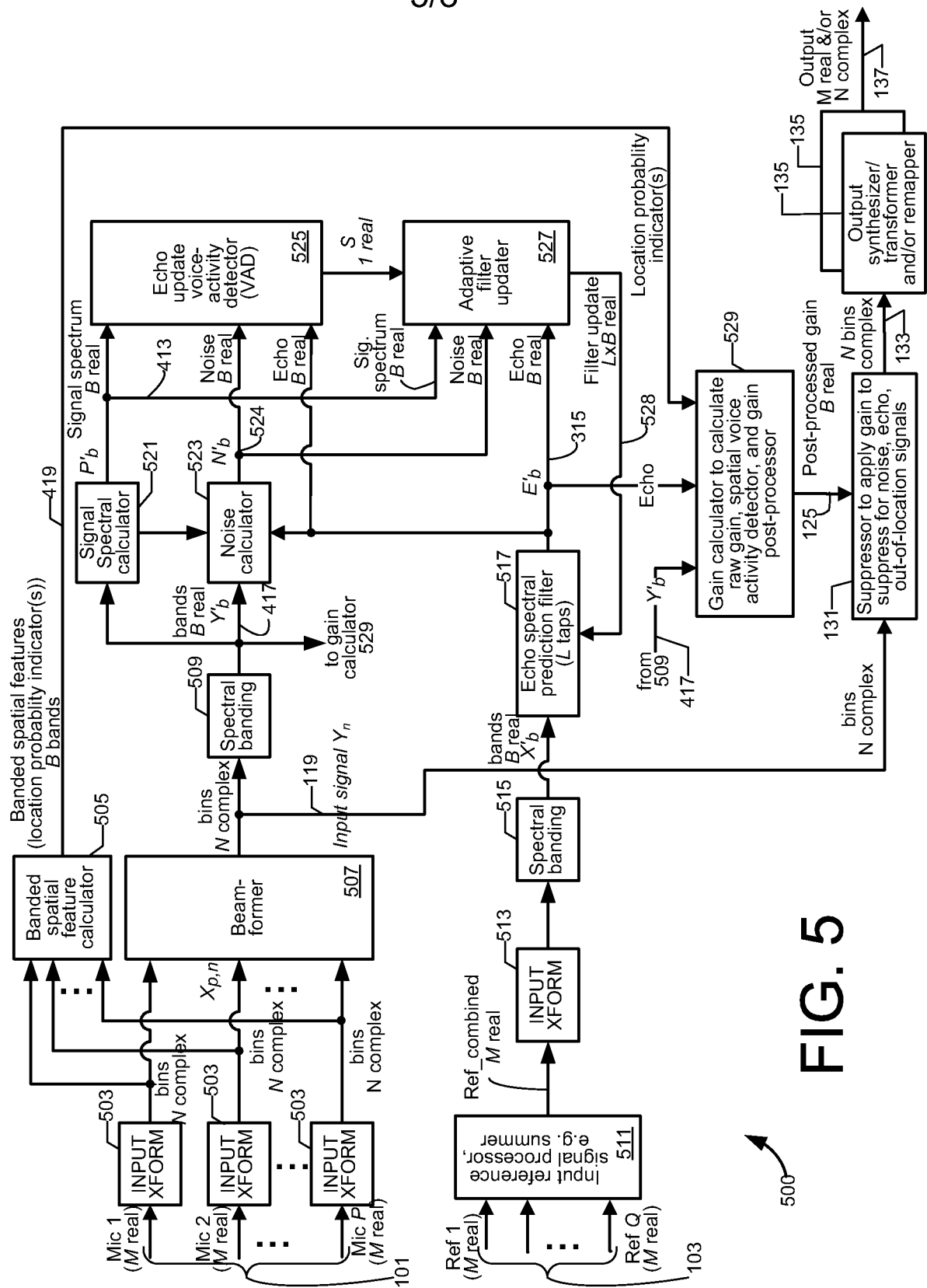


FIG. 5

6/8

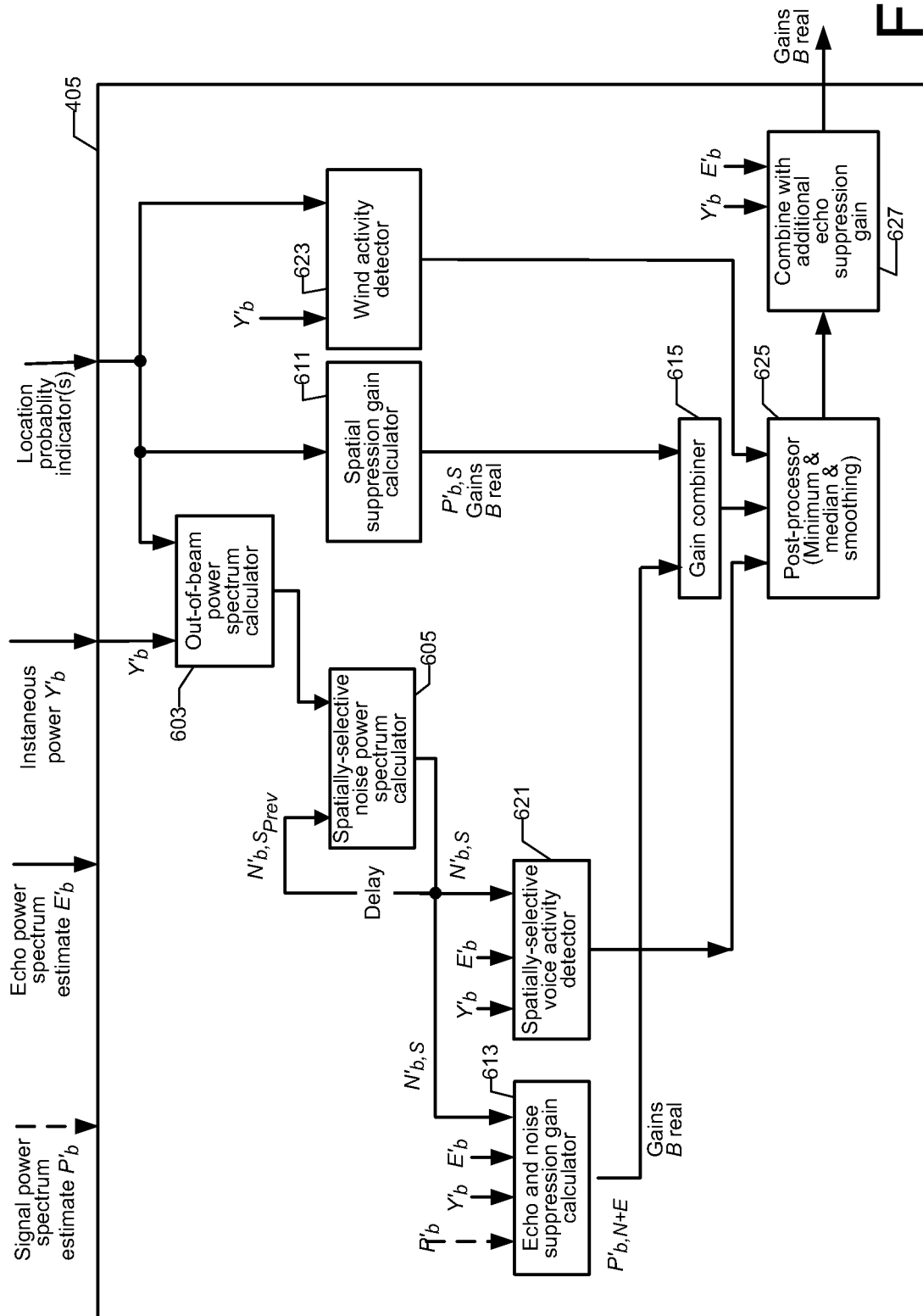


FIG. 6

7/8

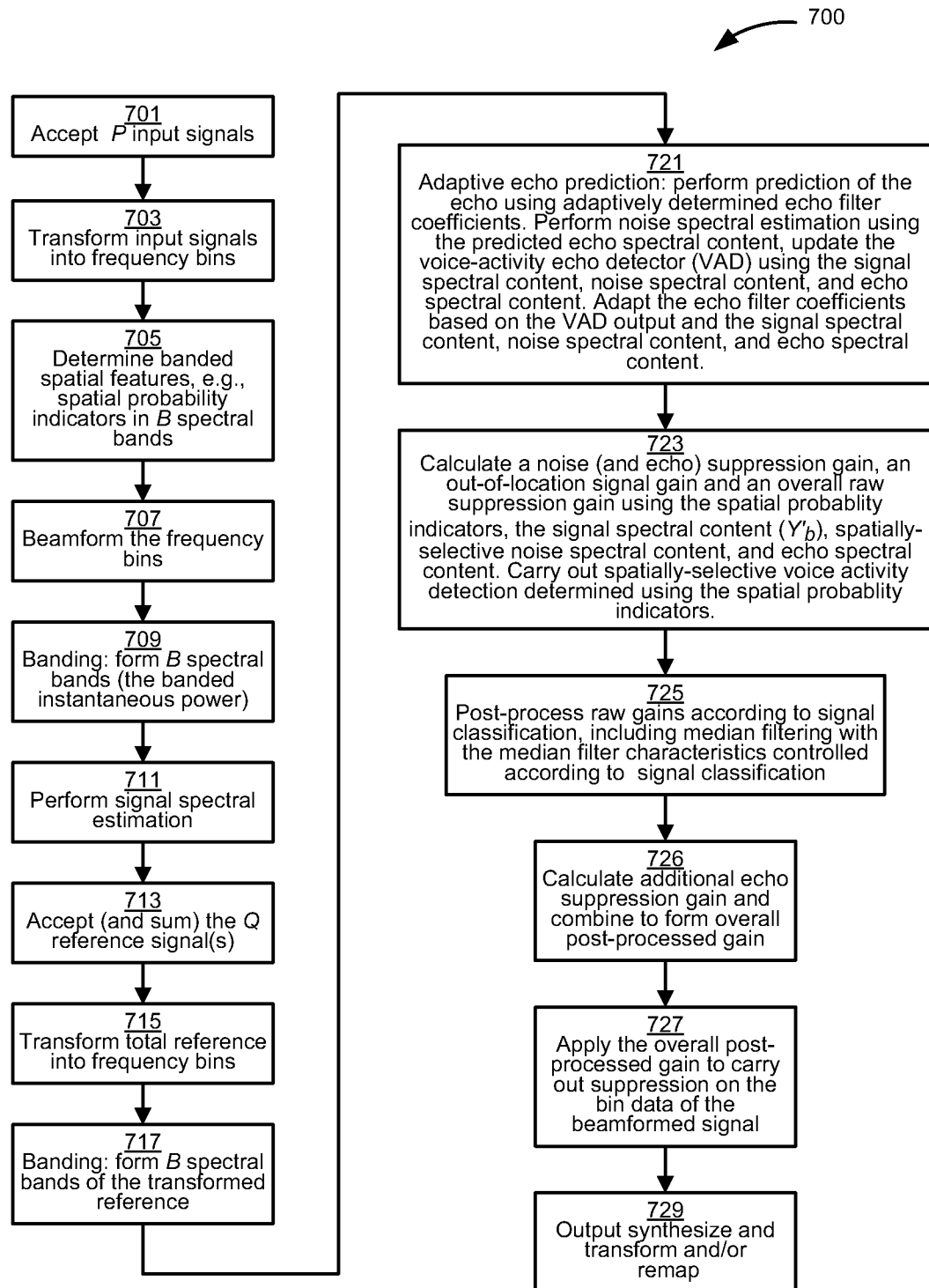
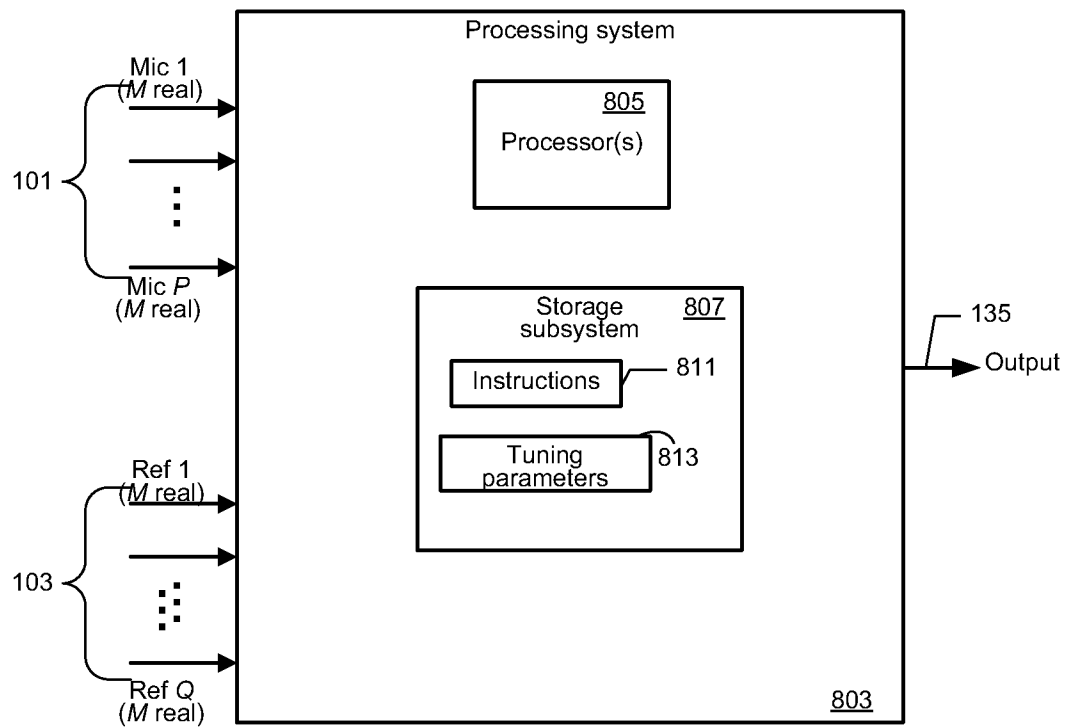


FIG. 7

8/8



800 ↗

FIG. 8

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2012/024372

A. CLASSIFICATION OF SUBJECT MATTER  
INV. G10L21/02  
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)  
G10L

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, INSPEC, COMPENDEX

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                         | Relevant to claim No.            |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| X         | DE 44 05 723 A1 (DAIMLER BENZ AG [DE])<br>24 August 1995 (1995-08-24)<br><br>column 1, line 1 - column 2, line 16<br>column 3, line 9 - column 3, line 23<br>column 4, line 58 - column 6, line 5<br>----- | 1-4,<br>9-19,<br>22-33,<br>38-42 |
| A         | EP 0 727 769 A2 (SONY CORP [JP])<br>21 August 1996 (1996-08-21)<br><br>page 3, line 45 - page 4, line 13<br>page 9, line 3 - page 10, line 29<br>-----                                                     | 1,2,<br>14-18,<br>30,31          |
| A         | US 2006/270467 A1 (SONG JIANMING J [US] ET<br>AL) 30 November 2006 (2006-11-30)<br><br>paragraph [0064] - paragraph [0065]<br>-----<br>-/-                                                                 | 1,2,<br>14-18,<br>30,31          |



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

24 May 2012

Date of mailing of the international search report

05/06/2012

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040,  
Fax: (+31-70) 340-3016

Authorized officer

De Meuleneire, M

## INTERNATIONAL SEARCH REPORT

International application No

PCT/US2012/024372

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                | Relevant to claim No.   |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------|-------------------------|
| A         | US 6 351 731 B1 (ANDERSON DAVID V [US] ET AL) 26 February 2002 (2002-02-26)<br><br>column 6, line 47 - column 7, line 11<br>----- | 1,2,<br>14-18,<br>30,31 |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2012/024372

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date    |
|-------------------------------------------|---------------------|----------------------------|------------------------|
| DE 4405723                                | A1                  | 24-08-1995                 | AT 185014 T 15-10-1999 |
|                                           |                     | DE 4405723 A1              | 24-08-1995             |
|                                           |                     | EP 0669606 A2              | 30-08-1995             |
|                                           |                     | ES 2138669 T3              | 16-01-2000             |
| -----                                     |                     |                            |                        |
| EP 0727769                                | A2                  | 21-08-1996                 | AT 209389 T 15-12-2001 |
|                                           |                     | AU 696187 B2               | 03-09-1998             |
|                                           |                     | AU 4444496 A               | 29-08-1996             |
|                                           |                     | BR 9600761 A               | 23-12-1997             |
|                                           |                     | CA 2169424 A1              | 18-08-1996             |
|                                           |                     | CN 1140869 A               | 22-01-1997             |
|                                           |                     | DE 69617069 D1             | 03-01-2002             |
|                                           |                     | DE 69617069 T2             | 11-07-2002             |
|                                           |                     | EP 0727769 A2              | 21-08-1996             |
|                                           |                     | ES 2163585 T3              | 01-02-2002             |
|                                           |                     | JP 3484801 B2              | 06-01-2004             |
|                                           |                     | JP 8221093 A               | 30-08-1996             |
|                                           |                     | PL 312845 A1               | 19-08-1996             |
|                                           |                     | RU 2127454 C1              | 10-03-1999             |
|                                           |                     | SG 52253 A1                | 28-09-1998             |
|                                           |                     | TR 960844 A2               | 21-10-1996             |
|                                           |                     | US 6032114 A               | 29-02-2000             |
| -----                                     |                     |                            |                        |
| US 2006270467                             | A1                  | 30-11-2006                 | NONE                   |
| -----                                     |                     |                            |                        |
| US 6351731                                | B1                  | 26-02-2002                 | NONE                   |
| -----                                     |                     |                            |                        |