(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
19 May 2016 (19.05.2016)

WIPO | PCT

(51) International Patent Classification:
H04B 15/00 (2006.01)

(21) International Application Number:
PCT/US2015/060337

(22) International Filing Date:
12 November 2015 (12.1 1.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/078,844   12 November 2014 (12. 11.2014)   US
14/938,816   11 November 2015 (11.11.2015)   US

(71) Applicant: CYPHER, LLC [US/US]; Suite 200, 282 E. 12200 South, Draper, Utah 84020 (US).

(72) Inventors: SHERWOOD, Erik; 282 E. 12200 South, Suite 200, Draper, Utah 84020 (US). GRUNDSTROM, Carl; 282 E. 12200 South, Suite 200, Draper, Utah 84020 (US).

(74) Agent: DORIUS, Kirk; Reed & Scardino LLP, Suite 1250, 301 Congress Ave., Austin, Texas 78701 (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, BN, 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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.1 7(H))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.1 7(in))*

(54) Title: ADAPTIVE INTERCHANNEL DISCRIMINITIVE RESCALING FILTER
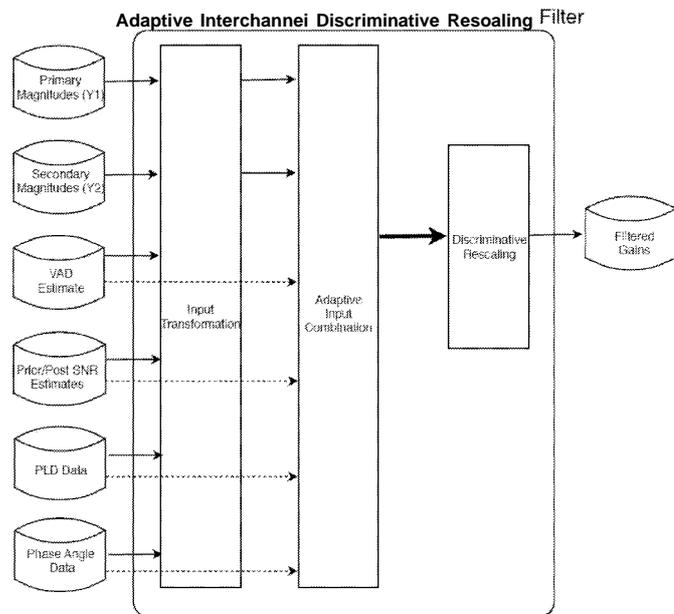
FIGURE 1

(57) Abstract: A method for filtering an audio signal includes modeling a probability density function (PDF) of a fast Fourier transform (FFT) coefficient of primary and reference channels; maximizing the PDFs to provide a discriminative relevance difference (DRD) between a noise magnitude estimate of the reference channel and a noise magnitude estimate of the primary channel. The primary channel is emphasized when the spectral magnitude of the primary channel is stronger than that of the reference channel; and is deemphasized when the spectral magnitude of the reference channel is stronger than that of the primary channel. A multiplicative rescaling factor is applied to a gain computed in a prior stage of a speech enhancement filter chain, and gain is directly applied when there is no prior stage.

ADAPTIVE INTERCHANNEL DISCRIMINITIVE RESCALING FILTER

CROSS-REFERENCE TO RELATED APPLICATION

[0001]   This patent application claims priority to U.S. provisional application serial number 62/078,844 filed November 12, 2014 and titled "Adaptive Interchannel Discriminative Rescaling Filter," which is incorporated herein in its entirety by reference.

TECHNICAL FIELD

[0002]   This disclosure relates generally to techniques for processing audio signals, including techniques for isolating voice data, removing noise from audio signals, or otherwise enhancing the audio signals prior to outputting the audio signals. Apparatuses and systems for processing audio signals are also disclosed.

BACKGROUND

[0003]   A variety of audio devices, including state of the art mobile telephones, include a primary microphone that is positioned and oriented to receive audio from an intended source, and a reference microphone that is positioned and oriented to receive background noise while receiving little or no audio from the intended source. In many usage scenarios, the reference

microphone provides an indicator of the amount of noise that is likely to be present in a primary channel of an audio signal obtained by the primary microphone. In particular, the relative spectral power levels, for a given frequency band, between the primary and reference channel may indicate whether that frequency band is dominated by noise or by signal in the primary channel. The primary channel audio in that frequency band may then be selectively suppressed or enhanced accordingly.

[0004] It is the case, however, that the probability of speech (respectively, noise) dominance in the primary channel, considered as a function of the unmodified relative spectral power levels between the primary and reference channels, may vary by frequency bin and may not be stationary over time. Thus, the use of a raw power ratios, fixed thresholds, and/or fixed rescaling factors in interchannel comparison-based filtering may well result in undesirable speech suppression and/or noise amplification in the primary channel audio.

[0005] Accordingly, improvements are sought in estimating the differences in noise-dominant/speech-dominant power levels between input channels, and in suppressing noise and enhancing speech presence in the primary input channel.

SUMMARY OF THE INVENTION

[0006] One aspect of the invention features, in some embodiments, a method for transforming an audio signal. The method includes obtaining a primary channel of an audio signal with a primary microphone of an audio device; obtaining a reference channel of the audio signal with a reference microphone of the audio device; estimating a spectral magnitude of the primary channel of the audio signal for a plurality of frequency bins; and estimating a spectral magnitude of the reference channel of the audio signal for a plurality of frequency bins. The method further includes transforming one or more of the spectral magnitudes for

one or more frequency bins by applying at least one of a fractional linear transformation and at higher order rational functional transformation; and further transforming one or more of the spectral magnitudes for one or more frequency bins. The further transformation can include one or more of: renormalizing one or more of the spectral magnitudes; exponentiating one or more of the spectral magnitudes; temporal smoothing of one or more of the spectral magnitudes; frequency smoothing of one or more of the spectral magnitudes; VAD-based smoothing of one or more of the spectral magnitudes; psychoacoustic smoothing of one or more of the spectral magnitudes; combining an estimate of a phase difference with one or more of the transformed spectral magnitudes; and combining a VAD-estimate with one or more of the transformed spectral magnitudes.

[0007] In some embodiments, the method includes updating at least one of the fractional linear transformation and the higher order rational functional transformation per bin based on augmentative inputs.

[0008] In some embodiments, the method includes combining at least one of an a priori SNR estimate and an a posteriori SNR estimate with one or more of the transformed spectral magnitudes.

[0009] In some embodiments, the method includes combining signal power level difference (SPLD) data with one or more of the transformed spectral magnitudes.

[0010] In some embodiments, the method includes calculating a corrected spectral magnitude of the reference channel based on a noise magnitude estimate and a noise power level difference (NPLD). In some embodiments, the method includes calculating a corrected spectral magnitude of the primary channel based on the noise magnitude estimate and the NPLD.

[0011] In some embodiments, the method includes at least one of replacing one or more of the spectral magnitudes by weighted averages taken across neighboring frequency bins within a frame and replacing one or more of the spectral magnitudes by weighted averages

taken across corresponding frequency bins from previous frames.

[0012]   Another aspect of the invention features, in some embodiments, a method for adjusting a degree of filtering applied to an audio signal. The method includes obtaining a primary channel of an audio signal with a primary microphone of an audio device; obtaining a reference channel of the audio signal with a reference microphone of the audio device; estimating a spectral magnitude of the primary channel of the audio signal; and estimating a spectral magnitude of the reference channel of the audio signal. The method further includes modeling a probability density function (PDF) of a fast Fourier transform (FFT) coefficient of the primary channel of the audio signal; modeling a probability density function (PDF) of a fast Fourier transform (FFT) coefficient of the reference channel of the audio signal; maximizing at least one of a single channel PDF and a joint channel PDF to provide a discriminative relevance difference (DRD) between a noise magnitude estimate of the reference channel and a noise magnitude estimate of the primary channel; and determining which of the spectral magnitudes is greater for a given frequency. The method further includes emphasizing the primary channel when the spectral magnitude of the primary channel is stronger than the spectral magnitude of the reference channel; deemphasizing the primary channel when the spectral magnitude of the reference channel is stronger than the spectral magnitude of the primary channel; and wherein the emphasizing and deemphasizing include computing a multiplicative rescaling factor and applying the multiplicative rescaling factor to a gain computed in a prior stage of a speech enhancement filter chain when there is a prior stage, and directly applying a gain when there is no prior stage.

[0013]   In some embodiments, the multiplicative rescaling factor is used as a gain.

[0014]   In some embodiments, the method includes including an augmentative input with each spectral frame of at least one of the primary and reference audio channels.

[0015]   In some embodiments, the augmentative input includes estimates of an a priori SNR and an a posteriori SNR in each bin of the spectral frame for the primary channel. In

some embodiments, the augmentative input includes estimates of the per-bin NPLD between corresponding bins of the spectral frames for the primary channel and the reference channel. In some embodiments, the augmentative input includes estimates of the per-bin SPLD between corresponding bins of the spectral frames for the primary channel and reference channel. In some embodiments, the augmentative input includes estimates of a per frame phase difference between the primary channel and the reference channel.

[0016]   Another aspect of the invention features, in some embodiments, an audio device, including a primary microphone for receiving an audio signal and for communicating a primary channel of the audio signal; a reference microphone for receiving the audio signal from a different perspective than the primary microphone and for communicating a reference channel of the audio signal; and at least one processing element for processing the audio signal to filter and or clarify the audio signal, the at least one processing element being configured to execute a program for effecting any of the methods described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

A more complete understanding of the present invention may be derived by referring to the detailed description when considered in connection with the Figures, and

[0017]   FIG. 1 illustrates an adaptive interchannel discriminative rescaling filter process according to one embodiment.

[0018]   FIG. 2 illustrates input transformations for use in adaptive interchannel discriminative rescaling filter process according to one embodiment.

[0019]   FIG. 3 illustrates a comparison of noise and speech power levels according to one embodiment.

[0020] FIG. 4 illustrates an estimation of noise noise and speech power level probability distribution functions according to one embodiment.

[0021] FIG. 5 illustrates a comparison of noise and speech power levels according to one embodiment.

[0022] FIG. 6 illustrates an estimation of noise and speech power level probability distribution functions according to one embodiment.

[0023] FIG. 7 illustrates comparison of noise and speech power levels with estimates of discriminative gain functions according to one embodiment.

[0024] FIG. 8 illustrates a computer architecture for analyzing digital audio data.

DETAILED DESCRIPTION

[0025] The following description is of exemplary embodiments of the invention only, and is not intended to limit the scope, applicability or configuration of the invention. Rather, the following description is intended to provide a convenient illustration for implementing various embodiments of the invention. As will become apparent, various changes may be made in the function and arrangement of the elements described in these embodiments without departing from the scope of the invention as set forth herein. Thus, the detailed description herein is presented for purposes of illustration only and not of limitation.

[0026] Reference in the specification to "one embodiment" or "an embodiment" is intended to indicate that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least an embodiment of the invention. The appearances of the phrase "in one embodiment" or "an embodiment" in various places in the specification are not necessarily all referring to the same embodiment.

[0027]   The present invention extends to methods, systems, and computer program products for analyzing digital data. The digital data analyzed may be, for example, in the form of digital audio files, digital video files, real time audio streams, and real time video streams, and the like. The present invention identifies patterns in a source of digital data and uses the identified patterns to analyze, classify, and filter the digital data, e.g., to isolate or enhance voice data. Particular embodiments of the present invention relate to digital audio. Embodiments are designed to perform non-destructive audio isolation and separation from any audio source.

[0028]   The purpose of the Adaptive Interchannel Discriminative Rescaling (AIDR) filter is to adjust the degree of filtering of the spectral representation of the input from the primary microphone, which is presumed to contain more power from the desired signal than power from noise, based on the relevance-adjusted relative power levels of the primary and reference spectra, Y_1 and Y_2, respectively. The input from the reference microphone is presumed to contain more relevance-adjusted power from confounding noise than from the desired signal.

[0029]   If it is detected that the secondary microphone input tends to contain more speech than the primary microphone input (e.g. the user is holding the phone in a reversed orientation), then the expectation regarding the relative magnitudes of Y_1 and Y_2 will also be reversed. Then in the following description, the roles of Y_1 and Y_2, etc., are simply interchanged, with the exception that the gain modifications may continue to be applied to Y_1.

[0030]   The logic of the AIDR filter, roughly speaking, is that for a given frequency, when the reference input is stronger than the primary input, then the corresponding spectral magnitude in the primary input represents more noise than signal and should be suppressed (or at least not accentuated). When the relative strengths of the reference and primary input are reversed, the corresponding spectral magnitude in the primary input represents more signal than noise and should be accentuated (or at least not suppressed).

[**0031**]   However, accurately determining whether a given spectral component of the primary input is in fact "stronger" than its counterpart in the reference channel, in a manner relevant for noise suppression/speech enhancement contexts, typically requires one or both of the primary and reference spectral inputs be algorithmically transformed to a suitable form. Following transformation, filtering and noise suppression is effected via discriminative rescaling of the spectral components of the primary input channel. This suppression/enhancement is typically achieved by computing a multiplicative rescaling factor to be applied to gains computed in prior stages of a speech enhancement filter chain, although the rescaling factors may also be used as gains themselves with appropriate choice of parameters.

1   Filter Inputs

[**0032**]   A diagrammatic overview of the multistage estimation and discrimination process of the AIDR filter is presented in FIG. 1. Time-domain signals y_1, y_2 from the primary and secondary (reference) microphones are presumed to to have been processed into equal length frames of samples upstream from the AIDR filter, $y_i(s, t)$, where $i \in \{1, 2\}$, $s = 0, 1, \ldots$ is the sample index within the frame and $t = 0, 1, \ldots$ is the frame index. These sample frames will have been further converted to the spectral domain via Fourier transform, so that $y_i \rightarrow Y_i$ with $Y_i(k,m)$ indicating the k-th discrete frequency component ("bin") of the m-th spectral frame, where $k = 1, 2, \ldots, K$ and $m = 0, 1, \ldots$. Note that K, the number of frequency bins per spectral frame, is typically determined according to the sampling rate in the time domain, e.g. 512 bins for a sampling rate of 16 kHz. $Y_1(k,m)$ and $Y_2(k,m)$ are considered *necessary inputs* to the AIDR filter.

[**0033**]   If the AIDR filter is incorporated into a speech enhancement filter chain following other processing components, *augmentative inputs* carrying additional information may accompany each spectral frame. Particular example inputs of interest (used in different filter

variants) include:

1. Estimates of the *a priori* SNR $\xi(k,m)$ and *a posteriori* SNR $\eta(k,m)$ in each bin of the spectral frame for the primary signal. These values will typically have been computed by a previous statistical filtering stage, e.g. MMSE, Power Level Difference (PLD), etc. These are vector inputs of the same length as Y_i.

2. Estimates of $\alpha_{NPLD}(k,m)$, the per-bin noise power level difference (NPLD) between corresponding bins of the spectral frames for the primary and secondary signals. These values will have been computed by the PLD Filter. These are vector inputs of the same length as Y_i.

3. Estimates of $\alpha_{SPLD}(k,m)$, the per-bin speech power level difference (SPLD) between corresponding bins of the spectral frames for the primary and secondary signals. These values will have been computed by the PLD Filter. These are vector inputs of the same length as Y_i.

4. Estimates of S_1 and/or S_2, the probabilities of speech presence in the primary and secondary signals, computed by a previous voice activity detection (VAD) stage. It is assumed that scalars $S_i \in [0, 1]$.

5. Estimates of $\Delta\phi(m)$, the phase angle separation between the spectra of the primary and reference inputs in the m-th frame, as provided by a suitable prior processing stage, e.g. PHAT (phase transform), GCC-PHAT (generalized cross-correlation with phase transform), etc.

2 Stage 1a: Input transformation

[0034] The necessary inputs Y_i are combined into a single vector for use in discriminative rescaling (stage 2), as will be described shortly. An expanded diagram of the input transformation and combination process of the AIDR filter is presented in FIG. 2.This combination process does not necessarily act on the magnitudes $Y_i(k,m)$ directly, rather the raw magnitudes may first be transformed into more suitable representations $r_i(k,m)$ which act, for example, to smooth out temporal and inter-frequency fluctuations or reweight/rescale the magnitudes in a frequency-dependent manner.

[0035] Prototypical transformations ("Stage 1 Preprocessing") include

1. Renormalization of the magnitudes, e.g.

$$\bar{Y}_i(k,m) = \frac{Y(k,m)}{\sum_{i=1}^{K} Y(k,m)}.$$

2. Raising of magnitudes to a power, i.e. $\bar{Y}_i(k,m) = Y_i(k,m)^{p_i}$. Note that p_i may be negative, may not necessarily be integer-valued, and p_1 may not equal p_2. One effect of such a transformation, for appropriately chosen p_i, could be to accentuate differences by raising spectral peaks and flattening spectral troughs within a given frame.

3. Replacement of the magnitudes by weighted averages taken across neighboring frequency bins within a frame. This transformation provides a local smoothing in frequency and can help reduce negative effects of musical noise that may have been introduced in prior processing steps which may have already edited the FFT magnitudes. As an example, the magnitude $Y(k,m)$ may be replaced by the weighted average of its

value and the values of magnitudes of the adjacent frequency bins, via

$$
\bar{Y}(k,m) = \begin{cases} \dfrac{\sum_{i=-1}^{1} w_k(i+1)Y(k+i,m)}{\sum_{j=1}^{3} w_k(j)} & \text{if } k = 2,\ldots,K-1 \\[2ex] \dfrac{\sum_{i=0}^{1} w_k(i+1)Y(k+i,m)}{\sum_{j=2}^{3} w_k(j)} & \text{if } k = 1 \\[2ex] \dfrac{\sum_{i=-1}^{0} w_k(i+1)Y(k+i,m)}{\sum_{j=1}^{2} w_k(j)} & \text{if } k = K \end{cases}
$$

where $w_k \equiv (1,2,1)$ is a vector of frequency bin weights. The subscript k is included for w to acknowledge the possibility that the weighting vector for the local average could be different for different frequencies, e.g. narrower for low frequencies, broader for high frequencies. The weighting vector need not be symmetric about the k-th (central) bin. For instance, it may be skewed to weight more heavily bins above (in both bin index and corresponding frequency) the central bin. This may be useful during voiced speech to place emphasis on bins near the fundamental frequency and its higher harmonics.

4. Replacement of the magnitudes by weighted averages taken across corresponding bins from previous frames. This transformation provides temporal smoothing within each frequency bin and can help reduce negative effects of musical noise that may have been introduced in prior processing steps that may have already edited the FFT magnitudes. Temporal smoothing may be implemented in various ways. For example:

   a) Simple weighted averaging:

$$
\bar{Y}(k,m) = \begin{cases} \dfrac{\sum_{i=-2}^{0} w_k(i+3)Y(k,m+i)}{\sum_{j=1}^{3} w_k(j)}, w = (1,2,3) & \text{if } m \geq 2 \\[2ex] \dfrac{\sum_{i=-1}^{0} w_k(i+2)Y(k,m+i)}{\sum_{j=2}^{3} w_k(j)} & \text{if } m = 1 \\[2ex] Y(k,m) & \text{if } m = 0 \end{cases}
$$

   b) Exponential smoothing:

$$
\bar{Y}(k,m) = \begin{cases} Y(k,m) & \text{if } m = 0 \\[2ex] \beta Y(k,m) + (1-\beta)\bar{Y}(k,m-1) & \text{if } m > 0 \end{cases}
$$

Here $\beta \in [0, 1]$ is a smoothing parameter which determines the relative weighting of bin magnitudes from the current frame relative to previous frames.

5. Exponential smoothing with VAD-based weighting: It can also be useful to perform temporal smoothing in which bin magnitudes from only those prior frames which do/do not contain speech information are included. This requires sufficiently accurate VAD information (augmentative input) computed by a prior signal processing stage. VAD information may be incorporated into exponential smoothing as follows:

a)
$$\bar{Y}(k,m) = \begin{cases} \hat{Y}(k,m) & \text{if } m = 0 \\ \beta Y(k,m) + (1-\beta)\bar{Y}(k,m^*) & \text{if } m > 0 \end{cases}$$

In this variant, $m^* < m$ is the index of most recent prior frame such that $S_i(m^*)$ is above (or below) a specified threshold indicating speech presence/absence.

b) Alternatively, the probability of speech presence may be used to modify the smoothing rate directly:
$$\bar{Y}(k,m) = \begin{cases} \hat{Y}(k,m) & \text{if } m = 0 \\ \beta(S_i)Y(k,m) + (1-\beta(S_i))\bar{Y}(k,m^*) & \text{if } m > 0 \end{cases}$$

In this variant, β is a function of S_i, e.g. a sigmoid function with parameters chosen such that as S_i moves below (resp. above) a given threshold, $\beta(S_i)$ approaches a fixed value β_a (resp. β_b).

6. Reweighting according to psychoacoustic importance: mel-frequency and ERB-scale weighting.

[0036]   Note that any and/or all of the above stages may be combined, or some stages may be omitted, with their respective parameters adjusted according to application (e.g. mel-scale reweighting used for automatic speech recognition but not mobile telephony).

3   Stage 1b: Adaptive input combination

[0037]   The final output of the input transformation stage for frame index m is designated as $\mathbf{u}(m)$. Note that $\mathbf{u}(m)$ is a vector having the same length K as Y_i, and $u(k,m)$ indicates the component of $\mathbf{u}$ associated with the k-th discrete frequency component of the m-th spectral frame. The computation of $\mathbf{u}(m)$ requires the modified necessary inputs r_1, r_2, and in general form this is accomplished by a vector-valued function $\mathbf{f}: \mathrm{R}^{2K} \to \mathrm{R}^{K}$, $\mathbf{f}(r_1(m), r_2(m)) = \mathbf{u}(m)$.

[0038]   In its simplest implementation, the per-bin action of $\mathbf{f}$ on $r_1(k,m)$, $r_2(k,m)$ may be expressed as a fractional linear transformation:

$$f_k(\bar{Y}_1(k,m), \bar{Y}_2(k,m)) = u(k,m) = \frac{A_k \bar{Y}_1(k,m) + B_k}{C_k \bar{Y}_2(k,m) + D_k}$$

[0039]   Without loss of generality, larger values of $u(k,m)$ may be presumed to indicate that in the kth frequency bin there is more power from the desired signal than from the confounding noise at time index m.

[0040]   More generally, the numerator and denominator of f_k may instead involve higher order rational expressions in $r_1(k,m)$, $r_2(k,m)$:

$$f_k(\bar{Y}_1(k,m), \bar{Y}_2(k,m)) = u(k,m) = \frac{\sum_{i=0}^{Q} A_{i,k}(\bar{Y}_1(k,m))^i}{\sum_{j=0}^{R} C_{j,k}(\bar{Y}_2(k,m))^j}$$

[0041]   Furthermore, any piecewise smooth transformation may be represented within any desired order of accuracy with this general representation (Chisholm approximant). In addition, the transformation parameters (A_k, B_k, C_k, D_k, or $A_{i,k}$, $C_{j,k}$ in these example) may vary by frequency bin. For example, it can be useful to use different parameters for bins in lower versus higher frequency bands in cases where the expected noise power characteristics are different in lower versus higher frequencies.

[0042]   In practice, the parameters of f_k are not fixed but rather are updated from frame to frame based on augmentative inputs, e.g.

$$B_k = B_k(\alpha_{NPLD}(k,m), \xi(k,m), \eta(k,m), S_1(m), \Delta\phi(m)), \tag{1}$$

$$D_k = D_k(\alpha_{NPLD}(k,m), S_1(m), \Delta\phi(m)) \tag{2}$$

or

$$A_{i,k} = A_{i,k}(\alpha_{NPLD}(k,m), \xi(k,m), \eta(k,m), S_1(m), \Delta\phi(m)), \tag{3}$$

$$C_{j,k} = C_{j,k}(\alpha_{NPLD}(k,m), S_1(m), \Delta\phi(m)) \tag{4}$$

and so forth.

[0043]   The adjustments to the raw inputs $Y_1(k,m)$, $Y_2(k,m)$ effect a per-bin transformation of raw spectral power estimates to quantities more relevant to the purpose of discriminating which components of the input $Y_1(k,m)$ are predominantly relevant to the desired signal. The transformations may act, for example, to rescale relative peaks and troughs in the primary and/or reference spectra, to smooth (or sharpen) spectral transients, and/or to correct for differences in orientation or spatial separation between the primary and reference microphones. As such factors may change over time, the relevant parameters of the transformation are typically updated once per frame while the AIDR filter is active.

4 Stage 2: Discriminative rescaling

[0044]   The aim of the second stage is to filter noise components from the primary signal by reducing those $Y_1(k,m)$ magnitudes which are estimated to contain more noise than desired speech. The output of stage 1, $u(m)$, serves as this estimate. If we take the output

of stage 2 to be a vector of multiplicative gains for each frequency component of $Y_1(m)$, then the kth gain should be small (close to 0) when $u(k,m)$ indicates a very low SNR and large (near 1, e.g. if gains are restricted to be non-constructive) if $u(k,m)$ indicates a very high SNR. For the intermediate cases, it is desirable for there to be a gradual transition between these extremes.

[0045]   Phrased generally, in the second step of the filter, the vector **u** is converted piecewise-smoothly into a vector **w** in such fashion that small values u_k are mapped to small values w_k and large values u_k are mapped to larger non-negative values w_k. Here k indicates frequency bin index. This transformation is achieved via the vector-valued function $\mathbf{g} : R^N \to R^N$ giving $\mathbf{g}(\mathbf{u}) = \mathbf{w}$. Element-wise, **g** is described by non-negative piecewise smooth functions $g_k : R \to R$. It may well be the case that $0 \le w_k \le B_k$, for some finite B_k, but **g** need neither be bounded nor non-negative. Each g_k should, however, be finite and non-negative over the plausible range of inputs u_k.

[0046]   A prototypical example of **g** features the simple sigmoid function

$$g_k(u(k,m)) = w_k = \alpha_k + \frac{\beta_k}{1 + \exp(-\delta_k u(k,m) - \epsilon_k)}$$

in each coordinate.

[0047]   The generalized logistic function is more flexible:

$$g_k(u(k,m)) = w_k = \alpha_k + \frac{\beta_k - \alpha_k}{(1 + \gamma_k \exp(-\delta_k(u(k,m) - \mu_k)))^{\nu_k}}$$

[0048]   The parameter α_k sets the minimum value for w_k. It is typically chosen to be a small positive value, e.g. 0.1, to avoid total suppression of $Y(k,m)$.

[0049]   The parameter β_k is the primary determinant of the maximum value for w_k, and it is generally set to 1, so that high SNR components are not modified by the filter. For some applications, however, β_k may be made slightly larger than 1. When the AIDR is used as a post-processing component in a larger filtering algorithm, for example, and prior filtering

stages tend to attenuate the primary signal (either globally or in particular frequency bands), then $\beta_k > 1$ may act to restore some speech components that were previously suppressed.

[0050]   The output of g_k in the transitional, intermediate range of $u(k,m)$ values is determined by parameters δ_k, ν_k, and μ_k which control the degree, abscissa, and ordinate of maximum slope.

[0051]   Initial values of these parameters are determined by examining the distribution of $u(k,m)$ values for a variety of speakers under a wide range of noise conditions and comparing the $u(k,m)$ values to the relative power levels of noise and speech. These distributions may vary substantially with mixing SNR and noise type; there is less variation between speakers. There are also clear differences between (psychoacoustic/frequency) bands. Examples of probability distributions for noise vs. speech power levels within various frequency bands are shown in FIG. 3-6.

[0052]   The empirical curves thus obtained are well-matched by the generalized logistic functions. The generalized logistic function provides the best fits, though the simple sigmoid is often adequate. FIG. 7 shows a basic sigmoid function and a generalized logistic function fit to empirical probability data. A single 'best' parameter set can be found by aggregating many speakers and noise types, or parameter sets may be adapted to specific speakers and noise types.

5   Additional notes

[0053]   For convenience, $u(k,m)$ may be substituted for $u(k,m)$ in the (generalized) logistic function of Stage 2. This has the effect of concentrating values that may range over several orders of magnitude into a much smaller interval. The same end result may be achieved without resort to taking logarithms of the function input, however, by rescaling and algebraic recombination of parameter values using logarithms.

[0054]   Parameter values in Stage 2 may adjust on a decision-directed basis, within fixed limits.

[0055]   The vector **w** may be used either as a standalone vector of multiplicative gains to be applied to the spectral magnitudes of the primary input, or it may be used a scaling and/or shifting factor for gains computed in prior filtering stages.

[0056]   When used a standalone filter, the AIDR filter provides basic noise suppression using the modified relative levels of spectral powers as an ad hoc estimate of a priori SNR and the sigmoidal function as a gain function.

[0057]   Embodiments of the present invention may also extend to computer program products for analyzing digital data. Such computer program products may be intended for executing computer-executable instructions upon computer processors in order to perform methods for analyzing digital data. Such computer program products may comprise computer-readable media which have computer-executable instructions encoded thereon wherein the computer-executable instructions, when executed upon suitable processors within suitable computer environments, perform methods of analyzing digital data as further described herein.

[0058]   Embodiments of the present invention may comprise or utilize a special purpose or general-purpose computer including computer hardware, such as, for example, one or more computer processors and data storage or system memory, as discussed in greater detail below. Embodiments within the scope of the present invention also include physical and other computer-readable media for carrying or storing computer-executable instructions and/or data structures. Such computer-readable media can be any available media that can be accessed by a general purpose or special purpose computer system. Computer-readable media that store computer-executable instructions are computer storage media. Computer-readable media that carry computer-executable instructions are transmission media. Thus, by way of example, and not limitation, embodiments of the invention can comprise at least

two, distinctly different kinds of computer-readable media: computer storage media and transmission media.

[0059] Computer storage media includes RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage or other magnetic storage devices, or any other physical medium which can be used to store desired program code means in the form of computer-executable instructions or data structures and which can be accessed by a general purpose or special purpose computer.

[0060] A "network" is defined as one or more data links that enable the transport of electronic data between computer systems and/or modules and/or other electronic devices. When information is transferred or provided over a network or another communications connection (either hardwired, wireless, or a combination of hardwired or wireless) to a computer, the computer properly views the connection as a transmission medium. Transmission media can include a network and/or data links which can be used to carry or transmit desired program code means in the form of computer-executable instructions and/or data structures which can be received or accessed by a general purpose or special purpose computer. Combinations of the above should also be included within the scope of computer-readable media.

[0061] Further, upon reaching various computer system components, program code means in the form of computer-executable instructions or data structures can be transferred automatically from transmission media to computer storage media (or vice versa). For example, computer-executable instructions or data structures received over a network or data link can be buffered in RAM within a network interface module (e.g., a "NIC"), and then eventually transferred to computer system RAM and/or to less volatile computer storage media at a computer system. Thus, it should be understood that computer storage media can be included in computer system components that also (or possibly primarily) make use of transmission media.

[0062]   Computer-executable instructions comprise, for example, instructions and data which, when executed at a processor, cause a general purpose computer, special purpose computer, or special purpose processing device to perform a certain function or group of functions. The computer executable instructions may be, for example, binaries which may be executed directly upon a processor, intermediate format instructions such as assembly language, or even higher level source code which may require compilation by a compiler targeted toward a particular machine or processor. Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the described features or acts described above. Rather, the described features and acts are disclosed as example forms of implementing the claims.

[0063]   Those skilled in the art will appreciate that the invention may be practiced in network computing environments with many types of computer system configurations, including, personal computers, desktop computers, laptop computers, message processors, hand-held devices, multi-processor systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, mobile telephones, PDAs, pagers, routers, switches, and the like. The invention may also be practiced in distributed system environments where local and remote computer systems, which are linked (either by hardwired data links, wireless data links, or by a combination of hardwired and wireless data links) through a network, both perform tasks. In a distributed system environment, program modules may be located in both local and remote memory storage devices.

[0064]   With reference to FIG. 8 an example computer architecture 600 is illustrated for analyzing digital audio data. Computer architecture 600, also referred to herein as a computer system 600, includes one or more computer processors 602 and data storage. Data storage may be memory 604 within the computing system 600 and may be volatile or non-volatile memory. Computing system 600 may also comprise a display 612 for display of data

or other information. Computing system 600 may also contain communication channels 608 that allow the computing system 600 to communicate with other computing systems, devices, or data sources over, for example, a network (such as perhaps the Internet 610). Computing system 600 may also comprise an input device, such as microphone 606, which allows a source of digital or analog data to be accessed. Such digital or analog data may, for example, be audio or video data. Digital or analog data may be in the form of real time streaming data, such as from a live microphone, or may be stored data accessed from data storage 614 which is accessible directly by the computing system 600 or may be more remotely accessed through communication channels 608 or via a network such as the Internet 610.

[0065]   Communication channels 608 are examples of transmission media. Transmission media typically embody computer-readable instructions, data structures, program modules, or other data in a modulated data signal such as a carrier wave or other transport mechanism and include any information-delivery media. By way of example, and not limitation, transmission media include wired media, such as wired networks and direct-wired connections, and wireless media such as acoustic, radio, infrared, and other wireless media. The term "computer-readable media" as used herein includes both computer storage media and transmission media.

[0066]   Embodiments within the scope of the present invention also include computer-readable media for carrying or having computer-executable instructions or data structures stored thereon. Such physical computer-readable media, termed "computer storage media," can be any available physical media that can be accessed by a general purpose or special purpose computer. By way of example, and not limitation, such computer-readable media can comprise physical storage and/or memory media such as RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage or other magnetic storage devices, or any other physical medium which can be used to store desired program code means in

the form of computer-executable instructions or data structures and which can be accessed by a general purpose or special purpose computer.

[0067]   Computer systems may be connected to one another over (or are part of) a network, such as, for example, a Local Area Network ("LAN"), a Wide Area Network ("WAN"), a Wireless Wide Area Network ("WWAN"), and even the Internet 110. Accordingly, each of the depicted computer systems as well as any other connected computer systems and their components, can create message related data and exchange message related data (e.g., Internet Protocol ("IP") datagrams and other higher layer protocols that utilize IP datagrams, such as, Transmission Control Protocol ("TCP"), Hypertext Transfer Protocol ("HTTP"), Simple Mail Transfer Protocol ("SMTP"), etc.) over the network.

[0068]   Other aspects, as well as features and advantages of various aspects, of the disclosed subject matter should be apparent to those of ordinary skill in the art through consideration of the disclosure provided above, the accompanying drawings and the appended claims.

[0069]   Although the foregoing disclosure provides many specifics, these should not be construed as limiting the scope of any of the ensuing claims. Other embodiments may be devised which do not depart from the scopes of the claims. Features from different embodiments may be employed in combination.

[0070]   Finally, while the present invention has been described above with reference to various exemplary embodiments, many changes, combinations and modifications may be made to the embodiments without departing from the scope of the present invention. For example, while the present invention has been described for use in speech detection, aspects of the invention may be readily applied to other audio, video, data detection schemes. Further, the various elements, components, and/or processes may be implemented in alternative ways. These alternatives can be suitably selected depending upon the particular application or in consideration of any number of factors associated with the implementation or operation of

the methods or system. In addition, the techniques described herein may be extended or modified for use with other types of applications and systems. These and other changes or modifications are intended to be included within the scope of the present invention.

CLAIMS

What is claimed:

1. A method for transforming an audio signal comprising:

   obtaining a primary channel of an audio signal with a primary microphone of an audio device;

   obtaining a reference channel of the audio signal with a reference microphone of the audio device;

   estimating a spectral magnitude of the primary channel of the audio signal for a plurality of frequency bins;

   estimating a spectral magnitude of the reference channel of the audio signal for a plurality of frequency bins;

   transforming one or more of the spectral magnitudes for one or more frequency bins by applying at least one of a fractional linear transformation and a higher order rational functional transformation; and

   transforming one or more of the spectral magnitudes for one or more frequency bins by one or more of:

   renormalizing one or more of the spectral magnitudes;

   exponentiating one or more of the spectral magnitudes;

   temporal smoothing of one or more of the spectral magnitudes;

   frequency smoothing of one or more of the spectral magnitudes;

   VAD-based smoothing of one or more of the spectral magnitudes;

   psychoacoustic smoothing of one or more of the spectral magnitudes;

   combining an estimate of a phase difference with one or more of the transformed spectral magnitudes; and

   combining a VAD-estimate with one or more of the transformed spectral magnitudes.

2. The method of claim 1, further comprising updating at least one of the fractional linear transformation and the higher order rational functional transformation per bin based on augmentative inputs.

3. The method of claim 1, further comprising combining at least one of an a priori SNR estimate and an a posteriori SNR estimate with one or more of the transformed spectral magnitudes.

4. The method of claim 1, further comprising combining signal power level difference (SPLD) data with one or more of the transformed spectral magnitudes.

5. The method of claim 1, further comprising calculating a corrected spectral magnitude of the reference channel based on a noise magnitude estimate and a noise power level difference (NPLD).

6. The method of claim 5, further comprising calculating a corrected spectral magnitude of the primary channel based on the noise magnitude estimate and the NPLD.

7. The method of claim 1, further comprising at least one of replacing one or more of the spectral magnitudes by weighted averages taken across neighboring frequency bins within a frame and replacing one or more of the spectral magnitudes by weighted averages taken across corresponding frequency bins from previous frames.

8.   A method for adjusting a degree of filtering applied to an audio signal comprising:

obtaining a primary channel of an audio signal with a primary microphone of an audio device;

obtaining a reference channel of the audio signal with a reference microphone of the audio device;

estimating a spectral magnitude of the primary channel of the audio signal;

estimating a spectral magnitude of the reference channel of the audio signal;

modeling a probability density function (PDF) of a fast Fourier transform (FFT) coefficient of the primary channel of the audio signal;

modeling a probability density function (PDF) of a fast Fourier transform (FFT) coefficient magnitude of the reference channel of the audio signal; and

maximizing at least one of a single-channel PDF and a joint channel PDF to provide a rate of the discriminative relevance difference (DRD) between a noise magnitude estimate of the reference channel and a noise magnitude estimate of the primary channel;

determining which of the spectral magnitudes is greater for a given frequency;

emphasizing the primary channel when the spectral magnitude of the primary channel is stronger than the spectral magnitude of the reference channel;

deemphasizing the primary channel when the spectral magnitude of the reference channel is stronger than the spectral magnitude of the primary channel; and

wherein the emphasizing and deemphasizing include computing a multiplicative rescaling factor and applying the multiplicative rescaling factor to a gain computed in a prior stage of a speech enhancement filter chain when there is a prior stage, and directly applying a gain when there is no prior stage.

9. The method of claim 8, wherein the multiplicative rescaling factor is used as a gain.

10. The method of claim 8, further comprising including an augmentative input with each spectral frame of at least one of the primary and reference audio channels.

11. The method of claim 10, wherein the augmentative input comprises estimates of an a priori SNR and an a posteriori SNR in each bin of the spectral frame for the primary channel.

12. The method of claim 10, wherein the augmentative input comprises estimates of the per-bin NPLD between corresponding bins of the spectral frames for the primary channel and the reference channel.

13. The method of claim 10, wherein the augmentative input comprises estimates of the per-bin SPLD between corresponding bins of the spectral frames for the primary channel and reference channel.

14. The method of claim 10, wherein the augmentative input comprises estimates of a per frame phase difference between the primary channel and the reference channel.

15. An audio device, comprising:

a primary microphone for receiving an audio signal and for communicating a primary channel of the audio signal;

a reference microphone for receiving the audio signal from a different perspective than the primary microphone and for communicating a reference channel of the audio signal;

at least one processing element for processing the audio signal to filter and or clarify the audio signal, the at least one processing element being configured to execute a program for effecting a method comprising:

obtaining a primary channel of an audio signal with a primary microphone of an audio device;

obtaining a reference channel of the audio signal with a reference microphone of the audio device;

estimating a spectral magnitude of the primary channel of the audio signal;

estimating a spectral magnitude of the reference channel of the audio signal;

modeling a probability density function (PDF) of a fast Fourier transform (FFT) coefficient of the primary channel of the audio signal;

modeling a probability density function (PDF) of a fast Fourier transform (FFT) coefficient of the reference channel of the audio signal;

maximizing at least one of a single-channel PDF and a joint channel PDF to provide a discriminative relevance difference (DRD) between a noise magnitude estimate of the reference channel and a noise magnitude estimate of the primary channel;

determining which of the spectral magnitudes is greater for a given frequency;

emphasizing the primary channel when the spectral magnitude of the primary channel is stronger than the spectral magnitude of the reference channel;

deemphasizing the primary channel when the spectral magnitude of the reference channel is stronger than the spectral magnitude of the primary channel; and

wherein the emphasizing and deemphasizing include computing a multiplicative rescaling factor and applying the multiplicative rescaling factor to a gain computed in a prior stage of a speech enhancement filter chain when there is a prior stage, and directly applying a gain when there is no prior stage.

16. An audio device, comprising:

a primary microphone for receiving an audio signal and for communicating a primary channel of the audio signal;

a reference microphone for receiving the audio signal from a different perspective than the primary microphone and for communicating a reference channel of the audio signal;

at least one processing element for processing the audio signal to filter and or clarify the audio signal, the at least one processing element being configured to execute a program for effecting a method comprising:

obtaining a primary channel of an audio signal with a primary microphone of an audio device;

obtaining a reference channel of the audio signal with a reference microphone of the audio device;

estimating a spectral magnitude of the primary channel of the audio signal for a plurality of frequency bins;

estimating a spectral magnitude of the reference channel of the audio signal for a plurality of frequency bins;

transforming one or more of the spectral magnitudes for one or more frequency bins by applying at least one of a fractional linear transformation and a higher order rational functional transformation; and

transforming one or more of the spectral magnitudes for one or more frequency bins by one or more of:

renormalizing one or more of the spectral magnitudes;

exponentiating one or more of the spectral magnitudes;

temporal smoothing of one or more of the spectral magnitudes;

frequency smoothing of one or more of the spectral magnitudes;

VAD-based smoothing of one or more of the spectral magnitudes;

psychoacoustic smoothing of one or more of the spectral magnitudes;

combining an estimate of a phase difference with one or more of the transformed spectral magnitudes; and

combining a VAD-estimate with one or more of the transformed spectral magnitudes.

FIGURE 1
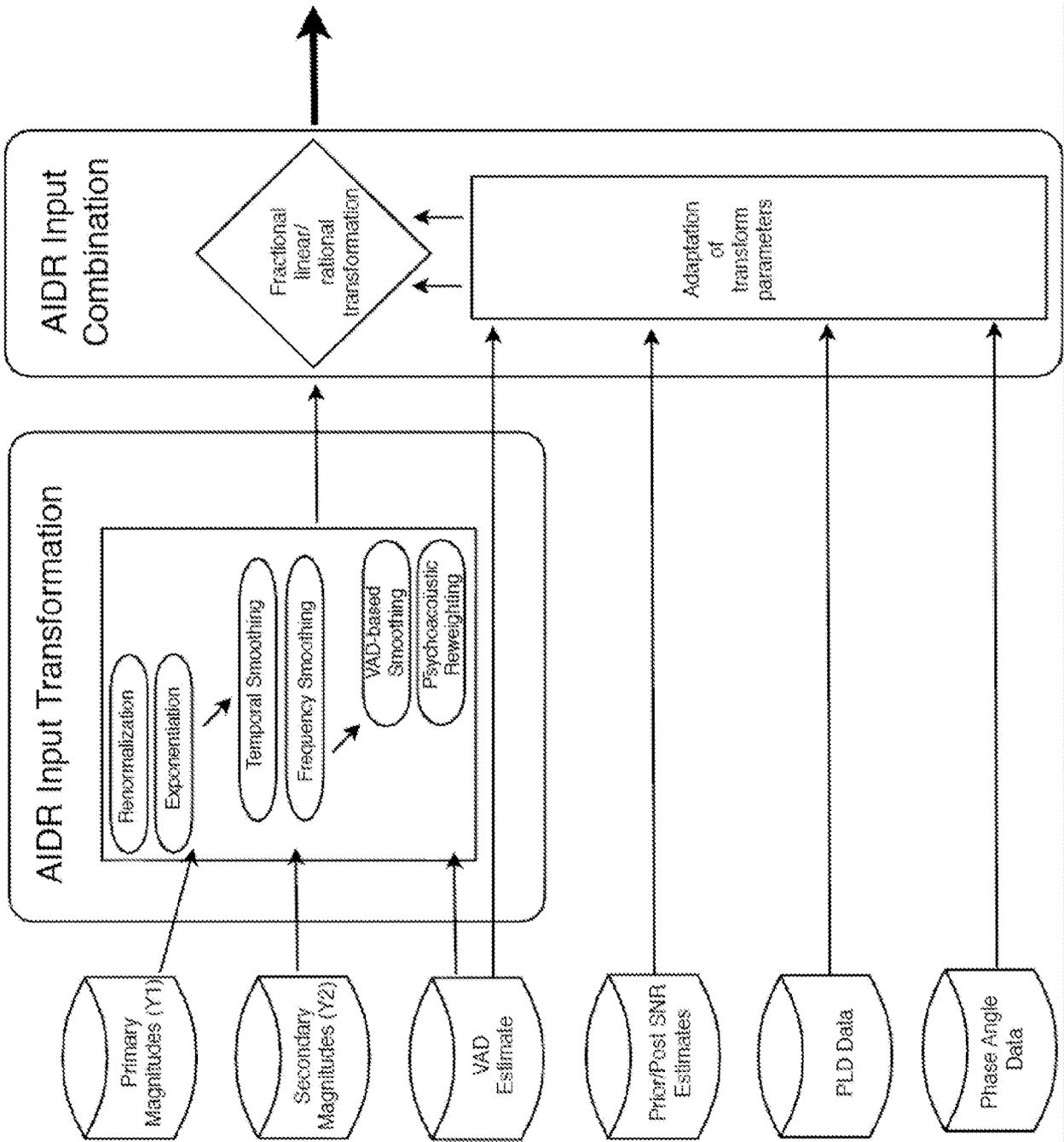
FIGURE 2
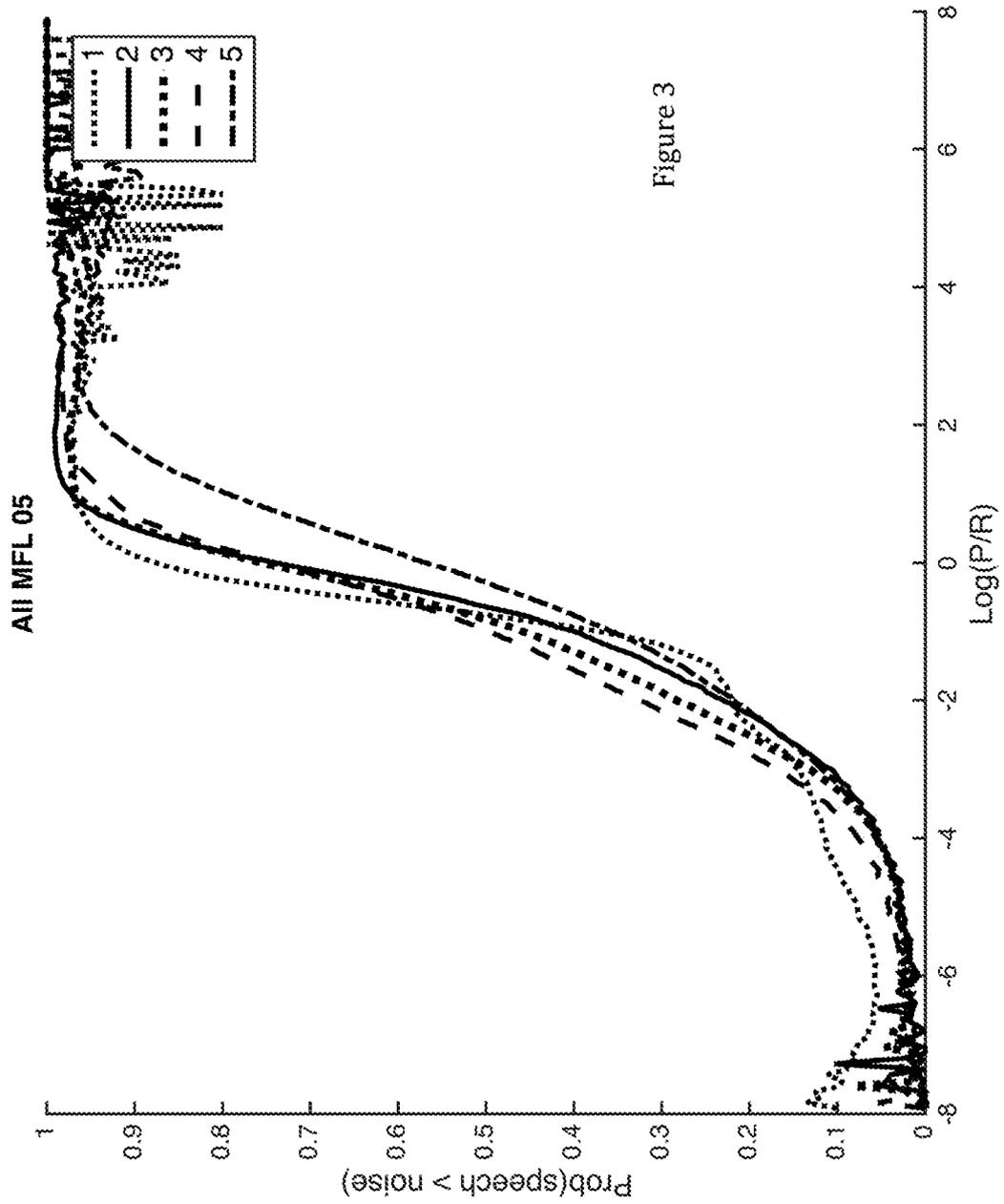
Figure 3
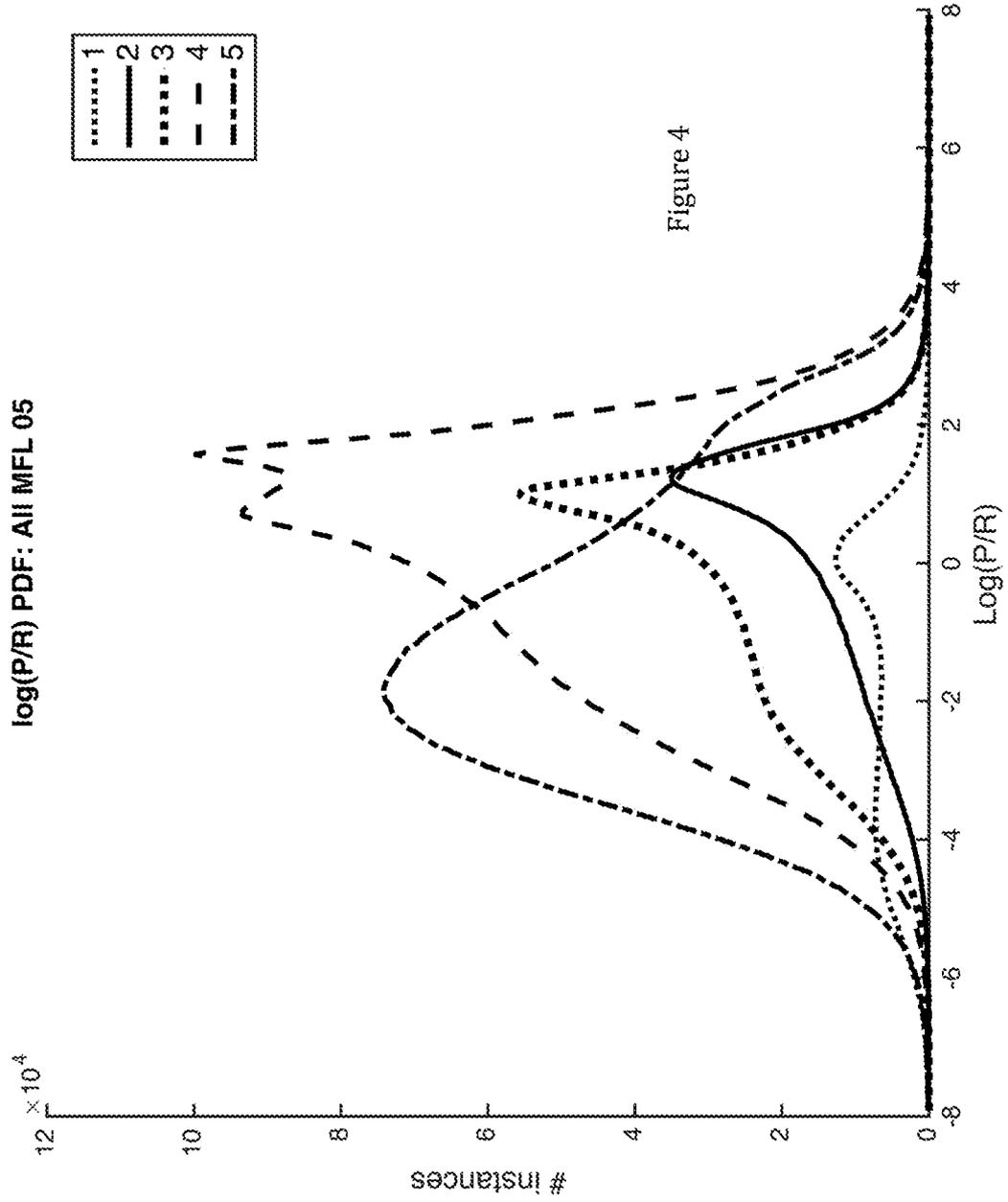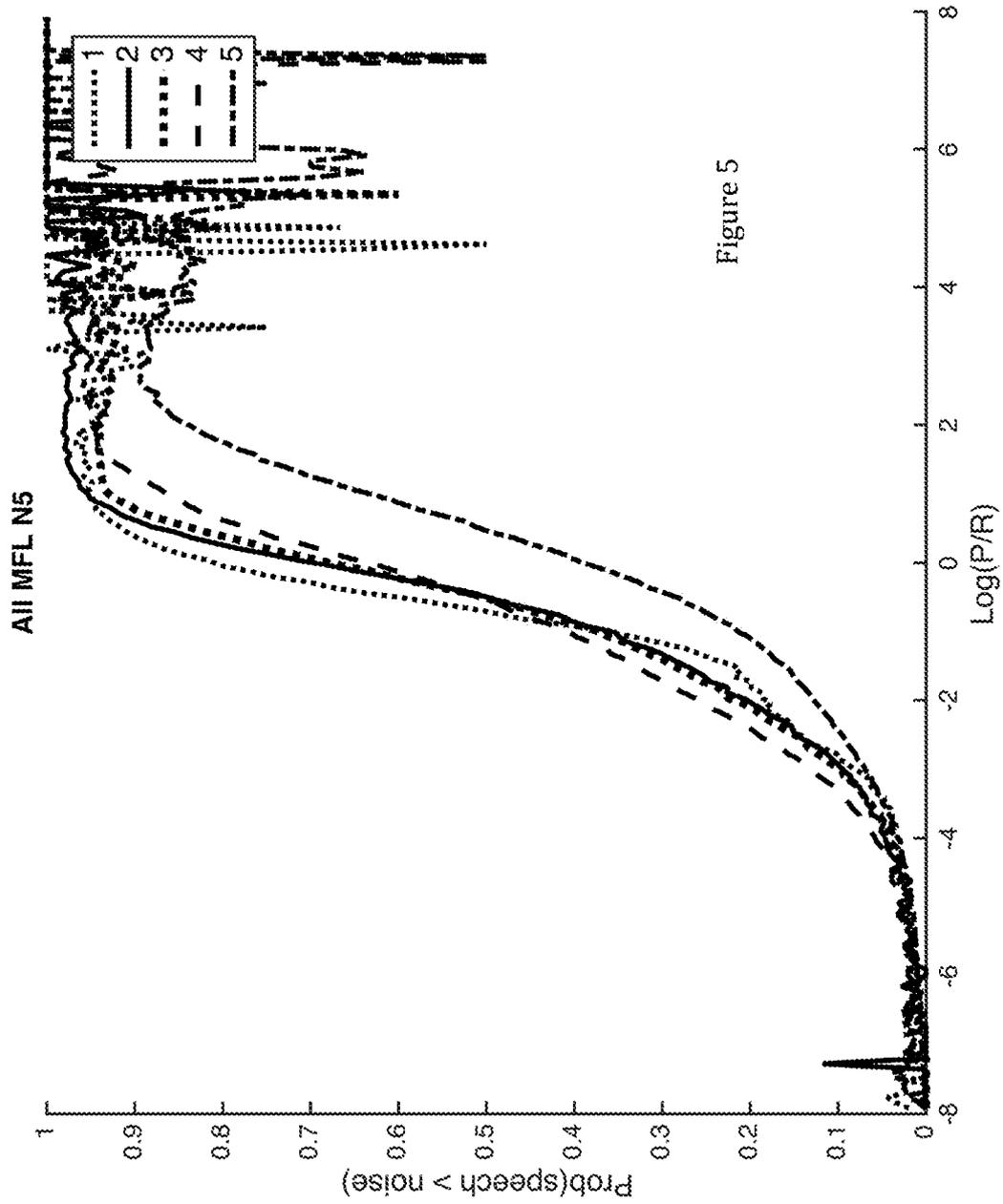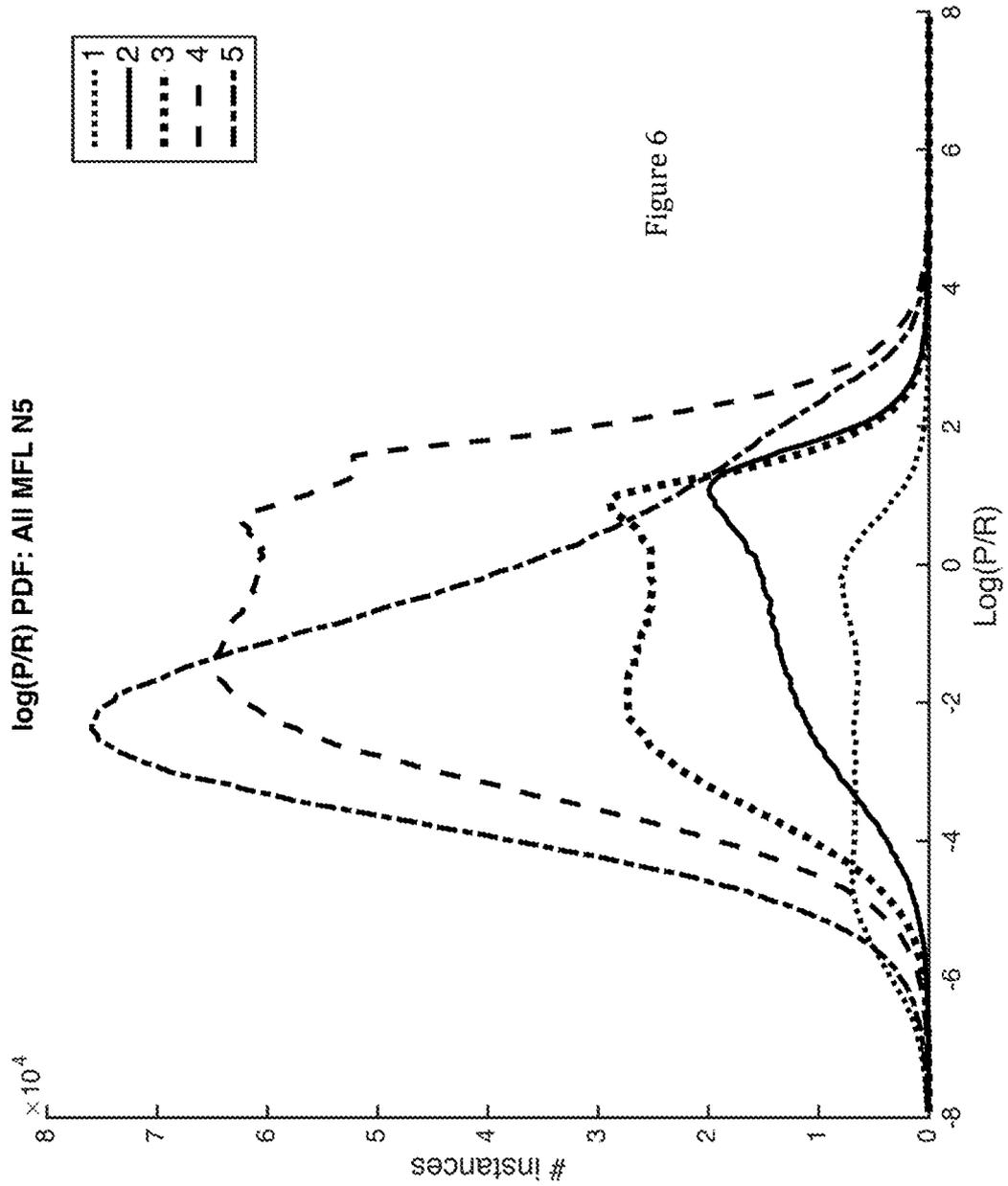
Figure 4
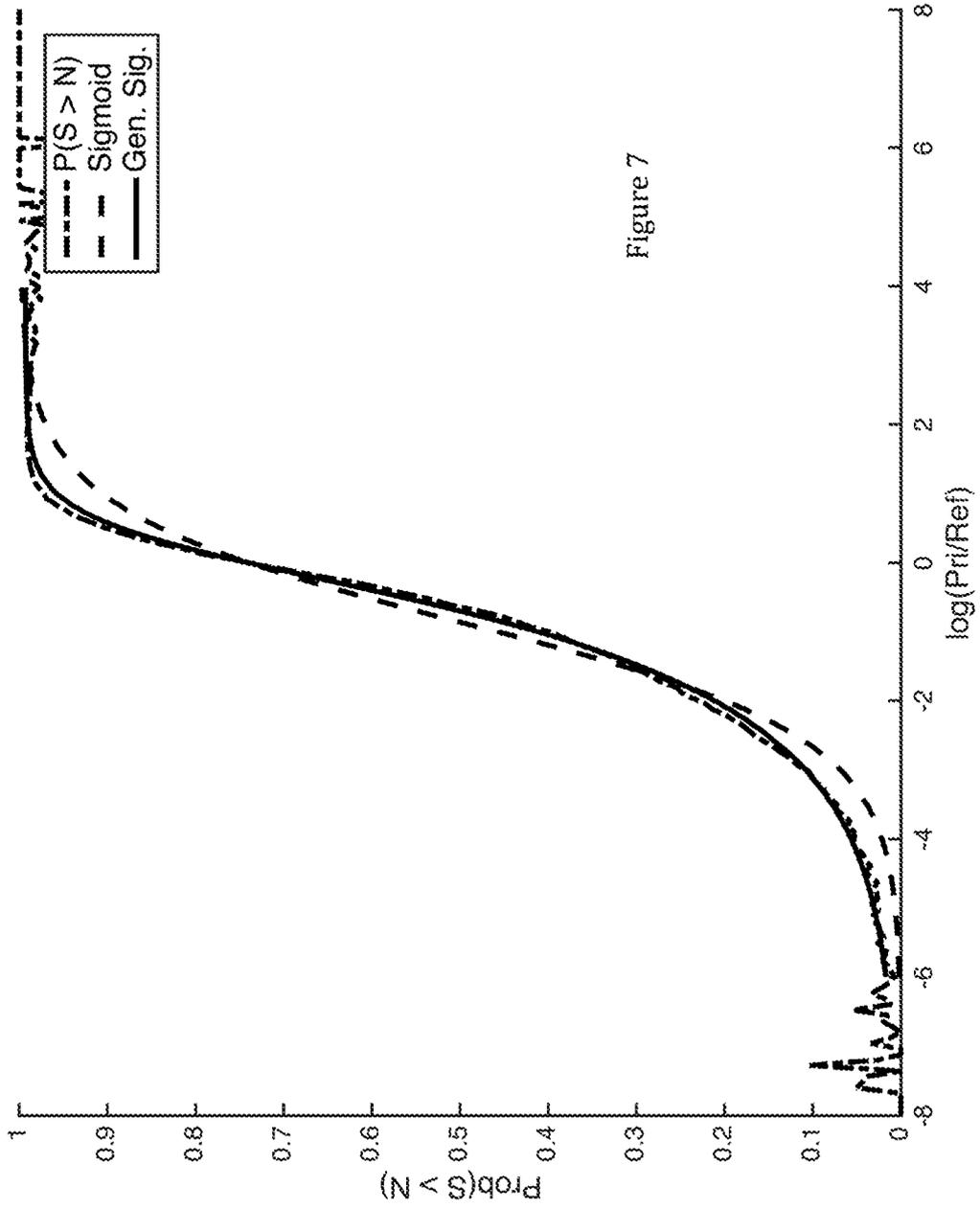
Figure 5
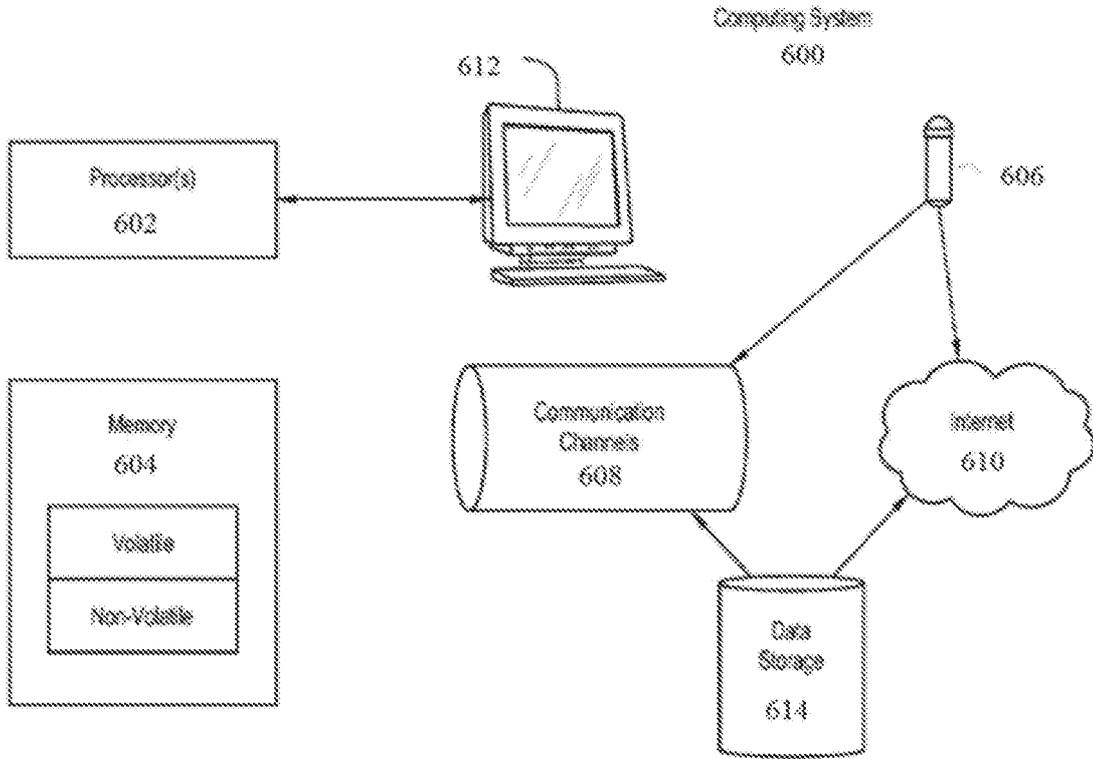
Figure 6

Figure 7

FIGURE 8

INTERNATIONAL SEARCH REPORT

	International application No.
	PCT/US2015/060337

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - H04B 15/00 (2015.01)
CPC - H04B 15/00 (2015.12)
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)
IPC(8) - G10L 21/02; G10L 25/84; H04B 15/00; H04B 1/3827 (2015.01)
CPC - G10L 21/02; G10L 25/84; H04B 15/00; H04B 1/3827 (2015.12)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
US: 381/122, 150; 704/226 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Orbit, Google Patents, Google Scholar, Google
Search terms used: Primary microphone, reference microphone, reference channel, estimating magnitude

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014/0029762 A1 (XIE et al) 30 January 2014 (30.01.2014) entire document	1-7, 16
Y	Pasha et al. Closed form Filtering for Linear Fractional Transformation Models. Proceedings of the 17th World Congress. IFAC, July 6-11, 2008. [retrieved on 2015-02-17]. Retrieved from the Internet:<URL: http://www.nt.ntnu.no/users/skoge/prost/proceedings/ifac2008/data/papers/2777.pdf >. entire document.	1-7, 16
Y	Lu et al. SPEECH ENHANCEMENT BY COMBINING STATISTICAL ESTIMATORS OF SPEECH AND NOISE. ICASSP 2010. [retrieved on 2015-02-17]. Retrieved from the Internet: <URL:http://www.mirlab.org/conference_papers/International_Conference/ICASSP%202010/pdfs/0004754.pdf>. entire document.	3
Y	US 2003/0206674 A1 (ALTMAN et al) 06 November 2003 (06.11.2003) entire document	4
Y	US 2013/0054231 A1 (JEUB et al) 28 February 2013 (28.02.2013) entire document	5, 6
Y	US 7,171,003 B1 (VENKATESH et al) 30 January 2007 (30.01.2007) entire document	7
A	US 20120226691 A1 (EDWARDS) 06 September 2012 (06.09.2012) entire document	1-7, 16

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

* Special categories of cited documents:	"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
"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	"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
"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)	"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
"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	"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 February 2016	**07 MAR 2016**

Name and mailing address of the ISA/	Authorized officer
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450	Blaine R. Copenheaver
Facsimile No. 571-273-8300	PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US20 15/060337

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

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

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

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

3. ☐ Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule $_6$.4(a).

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

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

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

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

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

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is
   restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
   1-7, 16

Remark on Protest   ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the
   payment of a protest fee.
   ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest
   fee was not paid within the time limit specified in the invitation.
   ☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 20 l5)

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-7,16, drawn to transforming an audio signal.
Group II, claims 8-15, drawn to adjusting a degree of filtering applied to an audio signal.

The inventions listed as Groups I-II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of the Group I invention: exponentiating one or more of the spectral magnitudes; temporal smoothing of one or more of the spectral magnitudes; frequency smoothing of one or more of the spectral magnitudes; VAD-based smoothing of one or more of the spectral magnitudes; psychoacoustic smoothing of one or more of the spectral magnitudes as claimed therein is not present in the invention of Group II. The special technical feature of the Group II invention: modeling a probability density function (PDF) of a fast Fourier transform (FFT) coefficient of the reference channel of the audio signal; maximizing at least one of a single channel PDF and a joint channel PDF to provide a discriminative relevance difference (DRD) between a noise magnitude estimate of the reference channel and a noise magnitude estimate of the primary channel as claimed therein is not present in the invention of Group I.

Groups I and II lack unity of invention because even though the inventions of these groups require the technical feature of obtaining a primary channel of an audio signal with a primary microphone of an audio device; obtaining a reference channel of the audio signal with a reference microphone of the audio device; estimating a spectral magnitude of the primary channel of the audio signal; estimating a spectral magnitude of the reference channel of the audio signal, this technical feature is not a special technical feature as it does not make a contribution over the prior art.

Specifically, US 2012/0226691 A1 (EDWARDS) 06 September 2012 (06.09.2012) teaches obtaining a primary channel of an audio signal with a primary microphone of an audio device (Para. 58); obtaining a reference channel of the audio signal with a reference microphone of the audio device (Patterns within the data, the data matching a pattern, or other data related to the received and interpreted data may be stored or referenced in one or more tables managed by the tables module 252, Para. 65); estimating a spectral magnitude of the primary channel of the audio signal (Plotting or graphically illustrating the three or more dimensions of the data may be used for spectral analysis, para. 106); estimating a spectral magnitude of the reference channel of the audio signal (Para. 102).

Since none of the special technical features of the Group I or II inventions are found in more than one of the inventions, unity of invention is lacking.