



(11) **EP 2 362 390 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention  
of the grant of the patent:  
**06.01.2016 Bulletin 2016/01**

(51) Int Cl.:  
**G10L 21/0208<sup>(2013.01)</sup> G10L 21/0232<sup>(2013.01)</sup>**

(21) Application number: **10250748.0**

(22) Date of filing: **09.04.2010**

(54) **Noise suppression**

Rauschunterdrückung

Suppression du bruit

(84) Designated Contracting States:  
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR  
HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL  
PT RO SE SI SK SM TR**

(30) Priority: **12.02.2010 IN DE03122010**

(43) Date of publication of application:  
**31.08.2011 Bulletin 2011/35**

(73) Proprietor: **NXP B.V.**  
**5656 AG Eindhoven (NL)**

(72) Inventors:  
• **Deshpande, Murali Mohan**  
**Redhill, Surrey RH1 1DL (GB)**  
• **Gundurao, Sudeendra Maddur**  
**Redhill, Surrey RH1 1DL (GB)**  
• **Goyens, Rob**  
**Redhill, Surrey RH1 1DL (GB)**  
• **Tirry, Wouter Joos**  
**Redhill, Surrey RH1 1DL (GB)**

• **Davies, Jeremy Thomas**  
**Redhill, Surrey RH1 1DL (GB)**

(74) Representative: **Crawford, Andrew et al**  
**NXP SEMICONDUCTORS UK LTD**  
**Intellectual Property & Licensing**  
**Red Central**  
**60 High Street**  
**Redhill, Surrey RH1 1SH (GB)**

(56) References cited:  
**US-A1- 2006 200 344 US-A1- 2007 192 102**

• **SZCZERBA M ET AL: "Pitch detection  
enhancement employing music prediction"**  
**JOURNAL OF INTELLIGENT INFORMATION**  
**SYSTEMS MARCH 2005 KLUWER ACADEMIC**  
**PUBLISHERS NL, vol. 24, no. 2-3, March 2005**  
**(2005-03), pages 223-251, XP002604517 DOI: DOI:**  
**10.1007/S10844-005-0324-6**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

## Description

**[0001]** The invention relates to detection and suppression or removal of noise in audio signals, with particular relevance for radio communication devices such as hand-portable radiotelephones.

**[0002]** Communication systems such as mobile phones are often used outdoors, which places strong requirements on the performance of any noise suppression systems present. One of the most dominant noise types is street noise, and in particular that caused by motorized traffic. Traffic noise can generally be classified into various types, but in many countries the sound of horns being used by cars and other vehicles such as autorickshaws strongly dominates the sound scene. Horn sounds in particular are perceived as annoying because they tend to be loud, and can adversely affect the quality and intelligibility of a conversation or can even inhibit communication altogether using mobile phones.

**[0003]** Current noise suppression algorithms can conveniently be divided into two categories, being single channel noise suppression systems and multi-channel noise suppression systems.

**[0004]** Most single-channel noise suppression systems are based on modification of the spectral amplitudes of an audio signal by means of a gain function. The calculation of a gain vector for implementing the gain function is carried out based on noise component estimation. A low SNR (signal to noise ratio) will significantly degrade the performance of such a noise estimator. In order to avoid speech degradation during low SNR situations, a conservative noise estimation process can be adopted, in which only stationary and slowly-varying non-stationary components are tracked by the noise estimator. This method can also be used independently of a speech detector. A faster noise component estimator to track non-stationary components would need to make use of a speech detector or even a dedicated speech model.

**[0005]** Multi-channel noise suppression systems can suppress both stationary and non-stationary noise. Most multi-channel noise suppressors rely on desired-speech detectors for calculating various parameters to obtain a noise reference. The gain vector is then evaluated based on the estimated noise.

**[0006]** Speech detectors or speech models are never 100% reliable, leading to speech degradation caused by imperfect gain calculations resulting from imperfect detector decisions. This problem is particularly prevalent under low SNR conditions. In addition, because the processing delay must be kept to a minimum in communication systems, only a short time window is available for deciding and estimating the noise. A time delay of longer than a few tens of milliseconds can have a noticeable impact on telephone conversations.

**[0007]** An example of a single channel noise suppression system 10 is illustrated in figure 1. An input audio input signal  $z(n)$  consists of the sum of a desired speech signal  $s(n)$  and a noise signal  $sn(n)$ . The audio input signal is sectioned into overlapping blocks and windowed. The windowed signal is transformed from the time domain into the frequency domain using an FFT (Fast Fourier Transform). These steps are represented by the windowing and FFT block 11 in figure 1. The magnitude spectrum obtained is then modified by a correction block 12 that applies a gain function in order to obtain an estimate of the speech signal,  $s(n)$ , which is then output by the system after inverse FFT and desectioning steps 12,13. The phase of the signal is left unchanged. The correction to the amplitude spectrum is obtained using a gain function that is determined for each frame and for each frequency bin in the frame.

**[0008]** Various existing methods are known for calculating the gain function based on different error criteria, for example as disclosed by R. Martin in "Spectral Subtraction Based on Minimum Statistics", Signal Processing VII: Theories and Application, pp. 1182-1185, EUSIPCO, 1994.

**[0009]** The gain vector obtained is used for modifying the real FFT of the audio signal frames according to the following relationship:

$$\hat{s}(i,k) = z_{ampl}(i,k) \cdot Gain(i,k) \quad (1)$$

where  $z_{ampl}(i,k)$  corresponds to the magnitude spectra of the input audio signal,  $Gain(i,k)$  is the gain factor and  $\hat{s}(i,k)$  is the modified magnitude spectrum of input audio signal. The index  $i$  corresponds to the frequency bin number and  $k$  corresponds to the frame number.

**[0010]** This modified amplitude spectrum is fed, together with the unmodified phase, to an IFFT (Inverse Fast Fourier Transform) block 13 for obtaining, after desectioning 14, the output signal  $\hat{s}(n)$  in which the noise  $sn(n)$  present in the input signal  $z(n)$  has been suppressed.

**[0011]** There are certain inherent limitations with existing noise-suppression algorithms, one of which relates to suppressing non-stationary noise such as horn sounds, as this requires steering the noise estimator to catch up with such noise using a speech detector. This approach has limitations because horn signals tend to have high energy and highly harmonic spectra that are normally detected incorrectly as speech, and under low SNR conditions the presence of such horn signals can cause speech detector performance to deteriorate. Furthermore, horn signals tend to occur only for very short durations (typically less than 1 second), so a noise estimator without a speech detector cannot normally be used effectively.

**[0012]** It is an object of the invention to address one or more of the above mentioned problems.

**[0013]** US 2006/0200344 discloses a method of reducing noise in an audio signal, in which a furrow filter is used to select spectral components that are narrow in frequency but relatively broad in time and a bar filter is used to select spectral components that are broad in frequency but relatively narrow in time, the relative energy distributions of the filters analysed to determine the optimal proportion of spectral components for an output signal.

**[0014]** US 2007/0192102 discloses a method of adaptively aligning windows to extract features according to the types and characteristics of voice signals, in which window lengths based on the window update points in a corresponding order are determined by employing the concept of a higher order peak and windows aligned according to window lengths.

**[0015]** M. Szczerba & A. Czyzewski, in "Pitch Detection Enhancement Employing Music Prediction", Journal of Intelligent Information Systems, 24:2/3, pp 223-251, 2005, disclose pitch detection methods widely used for extracting musical data from digital signals.

**[0016]** According to the invention there is provided a method of suppressing noise in an audio signal, as defined by the appended claims.

**[0017]** The method optionally comprises, for each of the sampled audio frames:

transforming the audio frame from the time domain into the frequency domain; and  
converting the resulting noise-suppressed audio frame back to the time domain,  
wherein the gain function comprises a gain vector applied in the frequency domain.

**[0018]** In alternative embodiments, the gain function may comprise a filter that is applied in the time domain, the filter being a notch filter having one or more notches at frequencies corresponding to the one or more detected spectral peaks.

**[0019]** The step of comparing a measure of high frequency content in the audio frame to a threshold value may comprise computing a sum of differences between consecutive samples, comparing the sum of differences to a second threshold value, and determining noise is present if the sum of differences exceeds the second threshold value

**[0020]** The step of comparing a measure of high frequency content in the audio frame to a threshold value may comprise computing a measure of energy in the audio frame, comparing this measure of energy to a third threshold value, and determining noise is present if the measure of energy exceeds the third threshold value.

**[0021]** The step of comparing a measure of high frequency content in the audio frame to a threshold value may comprise:

computing a sum of differences between consecutive samples;  
comparing the sum of differences to a second threshold value;  
computing a measure of energy in the audio frame;  
comparing the measure of energy to a third threshold value; and  
determining noise is present if the sum of the first and second numbers exceeds the first threshold value, the sum of differences exceeds the second threshold value and the measure of energy exceeds the third threshold value.

**[0022]** Generally, therefore, one or more of the above threshold values may be used in determining whether noise is present in each audio frame and, in a particular preferred embodiment, all three thresholds are used to determine whether noise is present.

**[0023]** Detecting a noise pattern in the sampled audio frame may be done by comparing frequencies of the spectrum of the audio frame with an average spectrum of the audio frame, a spectral peak being detected if a magnitude of a frequency exceeds the average spectrum by a preset factor.

**[0024]** The high frequency region of the audio signal spectrum is preferably a region over 2kHz. In preferred embodiments, this high frequency region will extend between 2kHz and half the frequency at which the audio signal is sampled.

**[0025]** The gain function may comprise a first gain function configured to emphasise a speech signal in the audio frame and a second gain function configured to suppress noise detected in the audio frame. The first gain function may be derived from a conventional speech detection process.

**[0026]** The audio signal in the method will typically comprise a speech signal and a noise signal, and the invention is particularly suited for when the noise signal is a vehicle horn noise.

**[0027]** The noise signal will generally be periodic and will have a harmonic structure, or in other words will comprise a fundamental frequency component and one or more harmonic components at other frequencies.

**[0028]** Embodiments according to the invention may be incorporated into a hand-portable radio communications device such as a radiotelephone that comprises a noise suppression module configured to perform the method of the invention.

**[0029]** The invention may also be embodied in a computer program for causing a computer to perform the method, which may be provided on a data carrier such as a memory chip, a computer-readable disc or other type of storage medium.

**[0030]** In a general aspect, the invention is based on using a noise signal detector and a filtering mechanism to suppress horn-like noise signals. The invention can be used together with single-channel or multi-channel noise suppression systems or as a standalone system for suppressing noise in the form of horn-like signals, thereby enhancing audio

intelligibility and quality.

**[0031]** Advantages of the invention relate to the detection of horn-like noise patterns instead of detection of speech. The detection of horn-like noise can be done more accurately than speech for low SNR situations, thereby making use of the invention more appropriate when an input audio signal is strongly affected by high energy high frequency non-stationary noise such as horn noises.

**[0032]** The detection of noise according to the invention operates on individual audio frames, and therefore operates effectively instantaneously. This type of detection can be used to steer, or modify, a noise suppression system that incorporates other noise suppression methods or used as a standalone noise removal system to specifically remove horn-like signals when detected.

**[0033]** The noise estimation part of an existing system could in practice be modified to adapt aggressively during presence of a horn signal. However, a generic solution would require a very reliable speech detector to avoid the problem of the noise component estimator being significantly biased by speech. Various methods have been tried in this direction but without success. Instead of trying to implement a robust speech detector in a noise suppression system that is also capable of handling horn-like signals, the invention provides a detector specifically directed to horn-like signals and uses this for suppression or removal of noise by spectral modification. The invention therefore offers a simpler solution to the problem of dealing with a particular type of noise that is likely to occur in practice.

**[0034]** Embodiments of the present invention will now be described by way of example and with reference to the accompanying drawings in which:

figure 1 is a schematic block diagram of a noise suppression system;  
figure 2 is a schematic block diagram of standalone horn noise suppression system;  
figure 3 is a schematic block diagram of a horn noise suppression system as part of a noise detection and suppression system;  
figures 4a and 4b are time and frequency domain representations of a single sampled audio frame of a rickshaw horn recording;  
figures 5a and 5b are time and frequency domain representations of a single sampled audio frame of a car horn recording;  
figures 6a and 6b are time and frequency domain representations of a single sampled audio frame of a truck horn recording;  
figures 7a and 7b are time and frequency domain representations of a single sampled audio frame of a motorcycle horn recording;  
figure 8 is a flow diagram illustrating operation of an exemplary embodiment of a time domain horn noise detector;  
figure 9 is a block diagram illustrating operation of a horn noise suppression system as part of a noise detection and suppression system; and  
figure 10 is a block diagram illustrating operation of a standalone horn noise suppression system.

**[0035]** Exemplary embodiments according to the invention comprise in general the following two steps:

1. Detection of a horn signal; and
2. Suppression or removal of the detected horn-like signal.

**[0036]** The horn detection and suppression system can be a standalone system or it could form part of a larger noise detection and suppression system. A basic block diagram of a standalone horn removal system 20 is shown in figure 2. A horn detection decision is made by a horn detector system 21, and a horn removal system 22 operates to suppress or remove the detected horn signal by applying a spectral gain function to the signal. The horn detector system 21 provides the input signal  $z$  to the horn removal system 22, together with an indication, provided in this case by a single bit horn detection flag, of whether horn noise has been detected in the frame in question.

**[0037]** A basic block diagram of a horn suppressor system provided as part of a larger noise detection and suppression system 30 is shown in figure 3. A noise suppression system 31 receives an input from a horn detector 21, which detects horn sounds on the input signal  $z$ . The noise suppression system 31 comprises a gain modification module 32 that is configured to compute a new gain for suppressing horn noise patterns whenever such horn sounds are detected. If no horn noise is detected, the gain modification module 31 suppresses noise in a conventional way, for example by the use of speech detection.

**[0038]** When designing a horn noise detector, it is necessary to understand the difference in characteristics between speech and horn-like sounds. Speech signals generally have the following characteristics:

- Limited zero-crossings in the time-domain signal, and with a signal energy concentrated in low frequency bands, at least for voiced sounds (i.e.  $<2$  kHz);

- Limited high-frequency transitions (typically < 20%);
- The energy of any high-frequency transitions is negligible;
- For unvoiced sounds, i.e. sounds other than vowel sounds that may have certain characteristics similar to noise (for example plosives such as 'b' and 'p' sounds made when the vocal folds are apart but are not vibrating), most energy is concentrated in the high frequency region. However, unvoiced speech signals have a low overall energy.

**[0039]** Horn-like signals, on the other hand, generally have the following characteristics:

- A high number of zero crossings and high energy in high frequency bands (> 2 kHz);
- High-frequency transitions occurring frequently (> 80%);
- Dominant high frequencies present;
- Energy of high-frequency transitions is considerable;
- Harmonic in nature, i.e. having a fundamental frequency and one or more harmonic over/undertone frequencies.

**[0040]** Figures 4 to 7 illustrate the main characteristics of typical horn-like signals. In figures 4a, 5a, 6a and 7a, the time-domain representation of audio recordings of rickshaw, car, truck and motorcycle horns respectively are shown, while in figures 4b, 5b, 6b and 7b the corresponding frequency domain representations of the same signals are shown. In each case, the audio frame comprises 80 samples, extending over a sample window of 10ms, i.e. corresponding to a sampling frequency of 8kHz, resulting in a maximum sampled frequency range of 0-4kHz. In each case, high frequency variations are visible as large variations in the values of alternate or consecutive samples. In most cases, the principle component of a horn noise will be augmented by other frequency components, though in some cases, as in the motorcycle horn noise in figures 7a and 7b, the principle frequency component at around 3200Hz dominates. In other cases such as the truck horn noise of figures 6a and 6b, multiple frequency component of roughly equal magnitude are present in addition to the principle component at 3200Hz.

**[0041]** Short duration audio frames such as these tend to have a poor frequency resolution when represented in the frequency domain. Detection of horn noise based on time domain analysis methods has however been found to be advantageous on frames of duration as short as 10ms.

#### High-frequency sample variations

**[0042]** Horn-like signals are highly varying and have harmonic spectra, i.e. generally comprise a fundamental frequency component together with harmonics at related frequencies. This characteristic can be used to detect such signals by determining the number of zero-crossing variations present in each frame. As used herein, the term 'zero crossings' refers to samples that fall either side of a zero line 41 (figure 4a). For a sampling frequency of 8kHz, the highest number of zero crossings will occur when sampling a 4kHz sine wave signal, where each sample alternates between a positive side of the signal and a negative side.

**[0043]** Two parameters related to zero-crossings can be used in detecting horn-like noise:

- First Order Consecutive Sample Variations (FOCSV), which are herein defined by consecutive samples that lead to a change in sign; and
- First Order Alternate Sample variations (FOASV), defined by alternate samples leading to a change in sign.

**[0044]** As an illustrative example, If  $x$  represents a frame of input audio samples and  $i$  represents a sample number in the frame, then the parameter FOCSV is computed as follows:

$$PrvDiff = x[i-1] - x[i-2]$$

$$CurDiff = x[i] - x[i-1]$$

$$FOCSV = \begin{cases} +1, & \text{if } ((PrvDiff == -ve) \& \& (CurDiff == +ve)) \\ +1, & \text{if } ((PrvDiff == +ve) \& \& (CurDiff == -ve)) \\ 0, & \text{otherwise} \end{cases}$$

[0045] In other words, the FOCSV parameter is determined to be 1 if both a previous pair and a current pair of samples involve a change in sign, and is zero otherwise.

[0046] The parameter FOASV, on the other hand is determined as follows:

$$Pr\ vDiff = x[i-2] - x[i-3]$$

$$CurDiff = x[i] - x[i-1]$$

$$FOASV = \begin{cases} +1, & \text{if } ((Pr\ vDiff = -ve) \& \& (CurDiff = +ve)) \\ +1, & \text{if } ((Pr\ vDiff = +ve) \& \& (CurDiff = -ve)) \\ 0, & \text{otherwise} \end{cases}$$

[0047] In other words, the FOASV parameter is determined to be 1 if two pairs of samples separated from each other by an intermediate sample involve a change in sign, and is zero otherwise.

[0048] In a frame containing N samples, the total number of high-frequency sample variations (*TotalHFVariations*) can be determined using the following relation:

$$TotalHFVariations = \sum_{i=0}^{N-2} FOCSV + \sum_{i=0}^{N-2} FOASV \quad (2)$$

where the terms FOCSV and FOASV are defined as above, and i is the sample number in each frame (which ranges from 0 to N-1). A frame can thereby be classified based on *TotalHFVariations* as being a horn or a non-horn frame. In practice, *TotalHFVariations* has been observed to be higher for frames having horn-like signals. A threshold (*Threshold<sub>HFV</sub>*) was determined experimentally considering a range of various signals. The following relationships can therefore be used to determine the presence of horn-like signals in each frame, based on this parameter:

$$\begin{aligned} TotalHFVariations &< Threshold_{HFV}, \text{ for non-horn signals} \\ TotalHFVariations &\geq Threshold_{HFV} \text{ for horn signals.} \end{aligned}$$

Energy of difference between consecutive samples

[0049] Horn-like signals also exhibit large amplitude differences between consecutive samples, which correspond to the signals having a high energy. The energy of the difference signal will therefore be comparatively higher for horn-like signals when compared to non-horn like signals. A term representing this energy can be based on a First Order Consecutive

[0050] Sample Difference (FOCSD) computed for various signal samples. This may be defined as follows:

$$FOCSD[i] = x[i] - x[i-1]$$

$$FOCSEnergy = \sum_{i=0}^{N-2} FOCSD[i]^2$$

[0051] In other words, the FOCSD energy parameter for a frame is determined from a sum of the squares of the differences between consecutive samples.

[0052] It has been observed that this FOCSD Energy will be higher for frames having horn-like signals than for frames having non horn-like signals. The following relations can be used to classify frames with horn and non-horn content:

$$FOCSEnergy < Threshold_{Energy}, \text{ for non-horn signals}$$

$$FOCSEnergy > Threshold_{Energy}, \text{ for horn signals.}$$

**[0053]** The threshold term,  $Threshold_{Energy}$ , was determined by analyzing the variations in FOCSEnergy in relation to the actual signal energy for various signals.

#### Instantaneous signal energy

**[0054]** Horn signals are generally non-stationary, occurring for only short durations (typically less than 4 seconds and often less than 1 second). Considering frame processing using blocks of 10ms each, horn signals can therefore span up to 400 frames. Horn signals have a high energy content throughout their duration. This property can be used to discriminate horn signals from unvoiced speech signals that may also have significant high-frequency content. The following relations can be used to classify frames with horn and non-horn content:

$$\begin{aligned} InstantaneousBlockEnergy &< Threshold_{AvEnergy}, \text{ for non-horn signals} \\ InstantaneousBlockEnergy &> Threshold_{AvEnergy}, \text{ for horn signals.} \end{aligned}$$

**[0055]**  $Threshold_{AvEnergy}$  can be determined by analyzing the variations of  $InstantaneousBlockEnergy$  in relation to an average signal energy for various un-voiced signals and horn-like signals.

**[0056]** The presence of a horn sound in a signal block is preferably decided based on all of the above three criteria. A flow diagram illustrating the process of determining whether noise is present in an audio frame is shown in figure 8. The process repeats (i.e. between points marked 'A') by operating on consecutive frames until there are no more frames to analyse, when the process ends (step 812).

**[0057]** As a first step 801, a time domain narrowband audio signal is sampled at 8kHz, resulting in successive blocks each consisting of N samples. For each block (step 802), three different tests are carried out. A first test involves computing the  $TotalHFVariations$  parameter, as detailed above (step 803), and comparing this parameter with a first threshold value  $Threshold_{HFV}$  (step 804). If the threshold value is not exceeded, the horn detection flag for that block is set to false (step 805), and the process continues to the next block (step 806).

**[0058]** A second test involves computing the FOCSEnergy parameter, as detailed above (step 807), and comparing this parameter with a second threshold value  $Threshold_{Energy}$  (step 808). If the threshold value is not exceeded, the horn detection flag for that block is set to false (step 805), and the process continues to the next block (step 806).

**[0059]** A third test involves computing the  $InstantaneousSignalEnergy$  parameter, as detailed above (step 809), and comparing this parameter with a third threshold value  $Threshold_{AvEnergy}$  (step 810). If the threshold value is not exceeded, the horn detection flag for that block is set to false (step 805), and the process continues to the next block (step 806).

**[0060]** Only if all three of the above threshold tests are passed does the process proceed to setting the horn detection flag for that block to true (step 811).

**[0061]** Following detection of horn noise in each frame, each frame is subjected to a noise suppression process. Depending on whether horn noise was detected in a frame, and whether the noise suppression system incorporates a conventional noise suppression process, the noise suppression process may i) leave the frame unchanged, ii) implement a conventional noise suppression process with no horn noise suppression, iii) implement horn noise suppression alone, or iv) implement both conventional noise suppression and horn noise suppression.

**[0062]** An exemplary noise detection and suppression system is illustrated in figure 9, in which a horn noise detection and suppression system is incorporated with a conventional speech-based noise suppression system. The characteristics of detected horn signals are incorporated into a modified gain vector to produce a modified magnitude spectrum that is adjusted to emphasise detected speech and to suppress any detected horn noises.

**[0063]** In a first step 901, an input signal  $z(n)$  is transformed into windowed FFT frames of size N, the value for N being chosen such that the signal can be considered to be stationary within each frame. A time domain input audio signal frame of N samples is thereby transformed into a frequency domain frame of  $N/2+1$  samples. The magnitude spectrum  $zamp(i)$  (step 902) is then used in the computation of a gain vector, and the phase part of the frame is neglected.

**[0064]** In step 903, assuming horn noise has been identified in a preceding time domain test (as described above), spectral peaks present in the magnitude spectrum  $zamp(i)$  are identified, which are taken to represent the horn signal present. This results in one or more indices of spectral peaks from the magnitude spectral bin values, which are used in a secondary gain computation step (step 907). The level for identifying the peaks is determined by calculating an average spectrum  $zamp_{avg}$ , given by the following equation:

$$zamp_{avg} = \frac{\sum_{i=0}^{N/2} zamp(i)}{(N/2 + 1)} \quad (3)$$

and the resulting magnitude spectral bin values are compared to  $zamp_{avg}$  multiplied by a peak detection factor  $\alpha$ , a

decision is made on whether to classify a particular sample from the magnitude spectrum as a peak value according to the following relationship:

$$\text{if}(z_{\text{ampl}}(i) \geq \alpha * z_{\text{ampl}}_{\text{avg}}), \{i \text{ is detected as peak index}\} \quad (3a)$$

**[0065]** The spectral bin indices satisfying the above relationship are identified as spectral peak bin numbers. The spectral peak indices identified are stored and used later for the gain computation by modifying gain vector when a horn sound is detected.

**[0066]** The noise floor used by noise suppression system for each frame is calculated by the Noise Floor Update block 903. The Noise Floor Estimate (NFE) is calculated by searching minima of the spectral bins over multiple frames and a noise floor used for each frame, i.e. Current Noise Floor (CNF) is updated in this block. The outputs of this block 904 are  $CNF(i)$  and  $NFE(i)$ . CNF is used in the subsequent gain calculation steps. An output from a speech detection block 905 is used by the Noise Floor Update block 904 for calculating NFE and CNF.

**[0067]** A gain computation block 906 receives  $CNF(i,k)$ ,  $z_{\text{ampl}}(i,k)$  and  $Y_N$  (gain factor) where  $i$  corresponds to a spectral bin number and  $k$  is the frame number. Computation of a gain,  $Gain_{ss}(i,k)$ , is given by the following relationship:

$$Gain_{ss}(i,k) = \frac{z_{\text{ampl}}(i,k) - Y_N \cdot CNF(i,k)}{z_{\text{ampl}}(i,k)} \quad (4)$$

**[0068]** In addition to the gain computation 906, a secondary gain computation block 907 is used to modify the gain computed in equation 4. In this block 907, a secondary gain vector is computed based on the previously defined horn noise detection information. The resulting secondary gain vector  $Gain_{sec}(i,k)$  is of size  $(N/2+1)$  and all the values in this vector are initialized to 1. This initialized value ensures that the gain computed by the gain computation block 906 is used when no horn-like signals have been detected in the present frame. The secondary gain computation block 907 block takes the horn detection flag and bin numbers calculated by the spectral peak detection block 903. The secondary gain  $Gain_{sec}(i,k)$  is calculated using the following relationship

$$Gain_{sec}(i,k) = 0 \quad (5)$$

where  $i$  is a spectral peak and corresponds to a frequency value of over 2000Hz. The secondary gain vector is used for modifying the gain calculated by the gain computation block 906, represented by combining block 908. All the elements of this vector are initialized to 1 in every frame before modification.

**[0069]** The resulting new gain vector,  $Gain_{new}(i,k)$ , that is then used for noise suppression is thereby computed using the following relationship:

$$Gain_{new}(i,k) = Gain_{sec}(i,k) * Gain_{ss}(i,k) \quad (6)$$

**[0070]** This new gain vector is applied to the real FFT data (in block 909), resulting in a modified magnitude spectrum (block 910). The modified spectrum 910 is passed through an inverse FFT block 912, resulting in a noise-suppressed signal. An equivalent operation is possible in the time domain, for example by applying a notch filter where the desired frequency response of the filter corresponds to the gain vector, i.e. where one or more notches in the filter correspond to the one or more spectral peaks that represent the noise that is to be suppressed.

**[0071]** Illustrated in figure 10 is a block diagram of a noise detection and suppression system 1000 in which a standalone horn suppression system is used, the main difference between this system and the system 900 in figure 9 being that a speech detection part of the system is not used. As before, an input signal  $z(n)$  is windowed and transformed to the frequency domain (block 1001), resulting in a magnitude spectrum (block 1002). A gain computation block 1004 takes in the spectral bin numbers from a spectral peak detection block 1003, the spectral bin numbers corresponding to spectral peaks identified in the magnitude spectrum. All the elements of the gain vector are initialized to 1 every frame before modification. On horn detection, the gain computation block 1004 computes a gain vector  $Gain(i,k)$  using the following relationship:

$$Gain(i,k) = 0 \quad (7)$$

where  $i$  is a spectral peak and corresponds to a frequency value of over 2000Hz.

**[0072]** This gain vector is then applied to the real FFT data (in combining block 1005), resulting in a modified magnitude spectrum (block 1006), to which the phase part 1007 is applied. The modified spectrum is passed through an inverse FFT block 1008 and a noise-suppressed signal is output.

**[0073]** Applications of embodiments of the invention described herein include speech enhancement devices used in communication/recording, audio enhancement during capture, editing and playback, and in audio scene analysis and steering of other processes such as noise adaptive audio or ringtone playback.

**[0074]** Other embodiments are also intentionally within the scope of the invention, which is defined by the following claims.

## Claims

1. A method of suppressing noise in an audio signal, the method comprising:

receiving an audio signal (801);  
dividing the received audio signal into a series of sampled audio frames (802); and  
for each of the sampled audio frames:

- i) determining whether noise is present in the audio frame by detecting a noise pattern in the sampled audio frame having one or more spectral peaks in a high frequency region of the audio signal spectrum; and
- ii) if noise is determined to be present in the audio frame (811), applying (908, 1005) a gain function to suppress the one or more spectral peaks in the sampled audio frame,

wherein the step of determining whether noise is present in the audio frame comprises comparing (803,804,807,808,809,810) a measure of high frequency content in the audio frame to a threshold value by computing (803) a first number of consecutive samples of opposite sign and a second number of alternate samples of opposite sign, comparing (804) a sum of the first and second numbers to a first threshold value, and determining (811, 805) noise is present if the sum exceeds the first threshold value.

2. The method of claim 1 comprising, for each of the sampled audio frames:

transforming (901, 1001) the audio frame from the time domain into the frequency domain; and  
converting (912, 1008) the resulting noise-suppressed audio frame back to the time domain,  
wherein the gain function comprises a gain vector applied in the frequency domain.

3. The method of claim 1 wherein the step of comparing a measure of high frequency content in the audio frame to a threshold value comprises computing (807) a sum of differences between consecutive samples, comparing (808) the sum of differences to a second threshold value, and determining (811, 805) noise is present if the sum of differences exceeds the second threshold value

4. The method of claims 1 or claim 3 wherein the step of comparing a measure of high frequency content in the audio frame to a threshold value comprises computing (809) a measure of energy in the audio frame, comparing (810) this measure of energy to a third threshold value, and determining (811, 805) noise is present if the measure of energy exceeds the third threshold value.

5. The method of claim 1 wherein the step of comparing a measure of high frequency content in the audio frame to a threshold value comprises:

computing (807) a sum of differences between consecutive samples;  
comparing (808) the sum of differences to a second threshold value;  
computing (809) a measure of energy in the audio frame;  
comparing (810) the measure of energy to a third threshold value; and  
determining (811, 805) noise is present if the sum of the first and second numbers exceeds the first threshold value, the third sum of differences exceeds the second threshold value and the measure of energy exceeds the third threshold value.

6. The method of any preceding claim wherein the step of determining whether noise is present in the audio frame is

carried out in the time domain.

7. The method of claim 6 wherein detecting a noise pattern in the sampled audio frame comprises comparing frequencies of the spectrum of the audio frame with an average spectrum of the audio frame, a spectral peak being detected if a magnitude of a frequency exceeds the average spectrum by a preset factor.
8. The method of claim any preceding claim wherein the high frequency region of the audio signal spectrum exceeds 2kHz.
9. The method of claim 1 or claim 2 wherein the gain function comprises a combination of a first gain function configured to emphasise a speech signal in the audio frame and a second gain function configured to suppress noise detected in the audio frame.
10. The method of any preceding claim wherein the noise signal has a harmonic structure.
11. The method of claim 10 wherein the noise signal is a vehicle horn noise.
12. A hand-portable radio communications device comprising a noise suppression module configured to perform the method of any one of claims 1 to 11.
13. A computer program for causing a computer to perform the method of any one of claims 1 to 11.

## Patentansprüche

1. Ein Verfahren zum Unterdrücken von Rauschen in einem Audiosignal, das Verfahren aufweisend:

Empfangen eines Audiosignales (801);  
Teilen des empfangenen Audiosignales in eine Serie von abgetasteten Audioframes (802); und  
für jeden der abgetasteten Audioframes:

- i) Bestimmen, ob Rauschen in dem Audioframe vorhanden ist mittels Detektierens eines Rauschmusters in dem abgetasteten Audioframe, welcher einen oder mehr spektrale Peaks in einem Hochfrequenz Bereich des Audiosignalspektrums hat; und
- ii) wenn bestimmt wurde, dass Rauschen in dem Audioframe (811) vorhanden ist, Anwenden (908, 1005) einer Verstärkungsfunktion, um den einen oder mehrere spektrale Peaks in dem abgetasteten Audioframe zu unterdrücken,

wobei der Schritt des Bestimmens, ob Rauschen in dem Audioframe vorhanden ist, aufweist

Vergleichen (803, 804, 807, 808, 809, 810) einer Messgröße von Hochfrequenzinhalt in dem Audioframe mit einem Schwellenwert mittels Berechnens (803) einer ersten Anzahl von fortlaufenden Abtastwerten von entgegengesetztem Vorzeichen und einer zweiten Anzahl von alternierenden Abtastwerten von entgegengesetztem Vorzeichen,  
Vergleichen (804) einer Summe der ersten und zweiten Anzahlen mit einem ersten Schwellenwert, und Bestimmen (811, 805) von Rauschen, welches vorhanden ist, wenn die Summe den ersten Schwellenwert übersteigt.

2. Das Verfahren gemäß Anspruch 1 aufweisend für jeden der abgetasteten Audioframes:

Transformieren (901, 1001) des Audioframes aus dem Zeitbereich in den Frequenzbereich; und Konvertieren (912, 1008) des resultierenden rauschunterdrückten Audioframes zurück in den Zeitbereich, wobei die Verstärkungsfunktion einen Verstärkungsvektor, welcher in dem Frequenzbereich angewandt wird, aufweist.

3. Das Verfahren gemäß Anspruch 1, wobei der Schritt des Vergleichens einer Messgröße von Hochfrequenzinhalt in dem Audioframe mit einem Schwellenwert aufweist  
Berechnen (807) einer Summe von Differenzen zwischen den fortlaufenden Abtastwerten,

Vergleichen (808) der Summe der Differenzen mit einem zweiten Schwellenwert, und Bestimmen (811,805) von Rauschen, welches vorhanden ist, wenn die Summe der Differenzen den zweiten Schwellenwert übersteigt.

4. Das Verfahren gemäß Anspruch 1 oder 3, wobei der Schritt des Vergleichens einer Messgröße von Hochfrequenzinhalt in dem Audioframe mit einem Schwellenwert aufweist  
Berechnen (809) einer Messgröße von Energie in dem Audioframe,  
Vergleichen (810) dieser Messgröße von Energie mit einem dritten Schwellenwert, und  
Bestimmen (811,805) von Rauschen, welches vorhanden ist, wenn die Messgröße der Energie den dritten Schwellenwert übersteigt.

5. Verfahren gemäß Anspruch 1, wobei der Schritt des Vergleichens einer Messgröße von Hochfrequenzinhalt in dem Audioframe mit einem Schwellenwert aufweist:

Berechnen (807) einer Summe von Differenzen zwischen fortlaufenden Abtastwerten;  
Vergleichen (808) der Summe von Differenzen mit einem zweiten Schwellenwert;  
Berechnen (809) einer Messgröße von Energie in dem Audioframe;  
Vergleichen (810) der Messgröße der Energie mit einem dritten Schwellenwert; und  
Bestimmen (811,805) von Rauschen, welches vorhanden ist, wenn die Summe der ersten und zweiten Anzahlen den dritten Schwellenwert übersteigt, die dritte Summe der Differenzen den zweiten Schwellenwert übersteigt und die Messgröße der Energie den dritten Schwellenwert übersteigt.

6. Das Verfahren gemäß einem der vorhergehenden Ansprüche, wobei der Schritt des Bestimmens, ob Rauschen in dem Audioframe vorhanden ist, in dem Zeitbereich ausgeführt wird.

7. Das Verfahren gemäß Anspruch 6, wobei Detektieren eines Rauschmusters in dem abgetasteten Audioframe Vergleichen von Frequenzen des Spektrums des Audioframes mit einem mittlerem Spektrum des Audioframes aufweist, wobei ein spektraler Peak detektiert wird, wenn eine Magnitude von einer Frequenz das mittlere Spektrum mittels eines voreingestellten Faktors übersteigt.

8. Das Verfahren gemäß einem der vorhergehenden Ansprüche, wobei der Hochfrequenzbereich des Audiosignalspektrums 2 kHz übersteigt.

9. Das Verfahren gemäß Anspruch 1 oder 2, wobei die Verstärkungsfunktion eine Kombination von einer ersten Verstärkungsfunktion, welche konfiguriert ist, um ein Sprachsignal in dem Audioframe hervorzuheben, und eine zweite Verstärkungsfunktion aufweist, welche konfiguriert ist Rauschen, detektiert in dem Audioframe, zu unterdrücken.

10. Das Verfahren gemäß einem der vorhergehenden Ansprüche, wobei das Rauschsignal eine harmonische Struktur hat.

11. Das Verfahren gemäß Anspruch 10, das Rauschsignal ein Fahrzeughupen Rauschen ist.

12. Eine Hand tragbare Funkkommunikation Vorrichtung aufweisend ein Rauschunterdrückungsmodul, welches konfiguriert ist, das Verfahren gemäß einem der Ansprüche 1 bis 11 auszuführen.

13. Ein Computerprogramm, welches einen Computer veranlasst das Verfahren gemäß einem der Ansprüche 1 bis 11 auszuführen.

## Revendications

1. Procédé de suppression du bruit dans un signal audio, le procédé consistant à:

recevoir un signal audio (801);  
diviser le signal audio reçu en une série de trames audio échantillonnées (802); et  
pour chacune des trames audio échantillonnées:

i) déterminer si du bruit est présent dans la trame audio en détectant un motif de bruit dans la trame audio échantillonnée, ayant un ou plusieurs pics spectraux dans une région à haute fréquence du spectre du

signal audio; et

ii) s'il est déterminé que du bruit est présent dans la trame audio (811), appliquer (908, 1005) une fonction de gain afin de supprimer les un ou plusieurs pics spectraux dans la trame audio échantillonnée,

dans lequel l'étape consistant à déterminer si du bruit est présent dans la trame audio consiste à comparer (803,804,807,808,809,810) une mesure de contenu à haute fréquence dans la trame audio à une valeur de seuil en calculant (803) un premier nombre d'échantillons consécutifs de signe opposé et un deuxième nombre d'échantillons alternés de signe opposé, à comparer (804) une somme des premier et deuxième nombres à une première valeur de seuil, et à déterminer (811, 805) que du bruit est présent si la somme dépasse la première valeur de seuil.

2. Procédé selon la revendication 1 consistant, pour chacune des trames audio échantillonnées, à:

transformer (901, 1001) la trame audio du domaine temporel au domaine fréquentiel; et reconvertir (912, 1008) la trame audio à bruit supprimé qui en résulte dans le domaine temporel, dans lequel la fonction de gain comprend un vecteur de gain appliqué dans le domaine fréquentiel.

3. Procédé selon la revendication 1, dans lequel l'étape consistant à comparer une mesure de contenu à haute fréquence dans la trame audio à une valeur de seuil consiste à calculer (807) une somme de différences entre des échantillons consécutifs, à comparer (808) la somme de différences à une deuxième valeur de seuil, et à déterminer (811, 805) que du bruit est présent si la somme de différences dépasse la deuxième valeur de seuil.

4. Procédé la revendication selon 1 ou la revendication 3, dans lequel l'étape consistant à comparer une mesure de contenu à haute fréquence dans la trame audio à une valeur de seuil consiste à calculer (809) une mesure de l'énergie dans la trame audio, à comparer (810) cette mesure de l'énergie à une troisième valeur de seuil, et à déterminer (811, 805) que du bruit est présent si la mesure de l'énergie dépasse la troisième valeur de seuil.

5. Procédé selon la revendication 1, dans lequel l'étape consistant à comparer une mesure de contenu à haute fréquence dans la trame audio à une valeur de seuil consiste à:

calculer (807) une somme de différences entre des échantillons consécutifs;  
comparer (808) la somme de différences à une deuxième valeur de seuil;  
calculer (809) une mesure de l'énergie dans la trame audio;  
comparer (810) la mesure de l'énergie à une troisième valeur de seuil; et  
déterminer (811, 805) que du bruit est présent si la somme des premier et deuxième nombres dépasse la première valeur de seuil, si la troisième somme de différences dépasse la deuxième valeur de seuil et si la mesure de l'énergie dépasse la troisième valeur de seuil.

6. Procédé selon l'une quelconque des revendications précédentes, dans lequel l'étape consistant à déterminer si du bruit est présent dans la trame audio est effectuée dans le domaine temporel.

7. Procédé selon la revendication 6, dans lequel la détection d'un motif de bruit dans la trame audio échantillonnée consiste à comparer des fréquences du spectre de la trame audio à un spectre moyen de la trame audio, un pic spectral étant détecté si un niveau d'une fréquence dépasse le spectre moyen d'un facteur prédéfini.

8. Procédé selon l'une quelconque des revendications précédentes, dans lequel la région à haute fréquence du spectre du signal audio dépasse 2 kHz.

9. Procédé selon la revendication 1 ou la revendication 2, dans lequel la fonction de gain comprend une combinaison d'une première fonction de gain configurée pour accentuer un signal vocal dans la trame audio et d'une deuxième fonction de gain configurée pour supprimer du bruit détecté dans la trame audio.

10. Procédé selon l'une quelconque des revendications précédentes, dans lequel le signal de bruit présente une structure harmonique.

11. Procédé selon la revendication 10, dans lequel le signal de bruit est un bruit de klaxon de véhicule.

12. Dispositif de radiocommunication portatif comprenant un module de suppression du bruit configuré pour mettre en

## EP 2 362 390 B1

oeuvre le procédé selon l'une quelconque des revendications 1 à 11.

13. Programme d'ordinateur destiné à amener un ordinateur à mettre en oeuvre le procédé selon l'une quelconque des revendications 1 à 11.

5

10

15

20

25

30

35

40

45

50

55

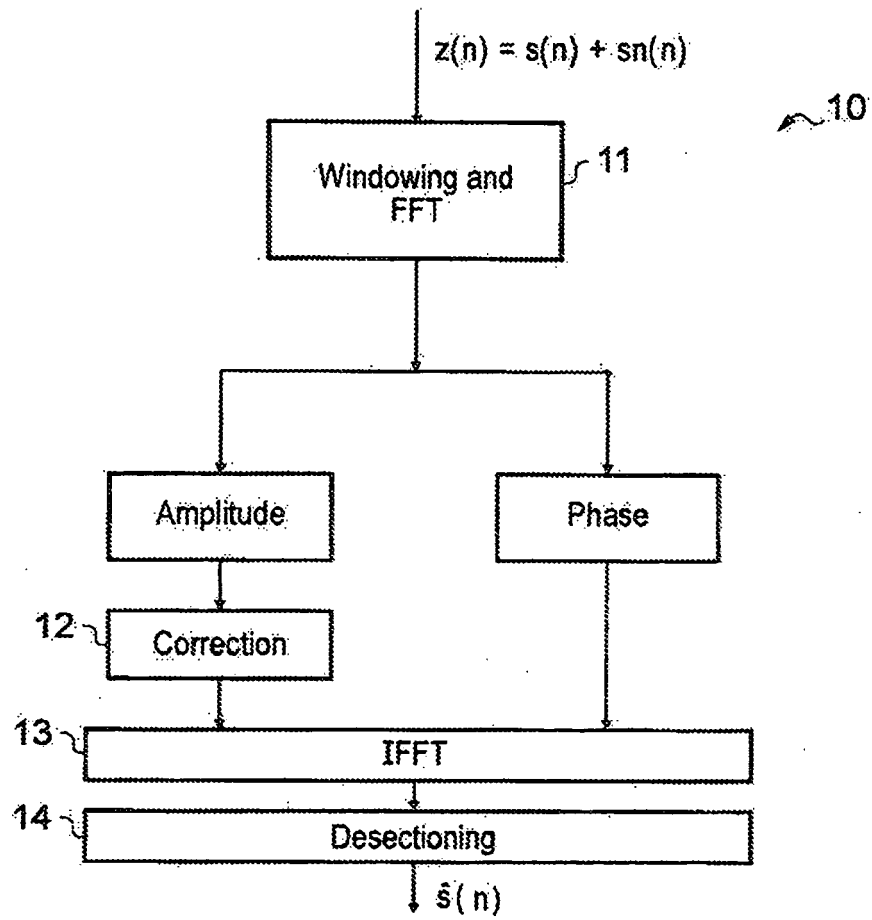


FIG. 1

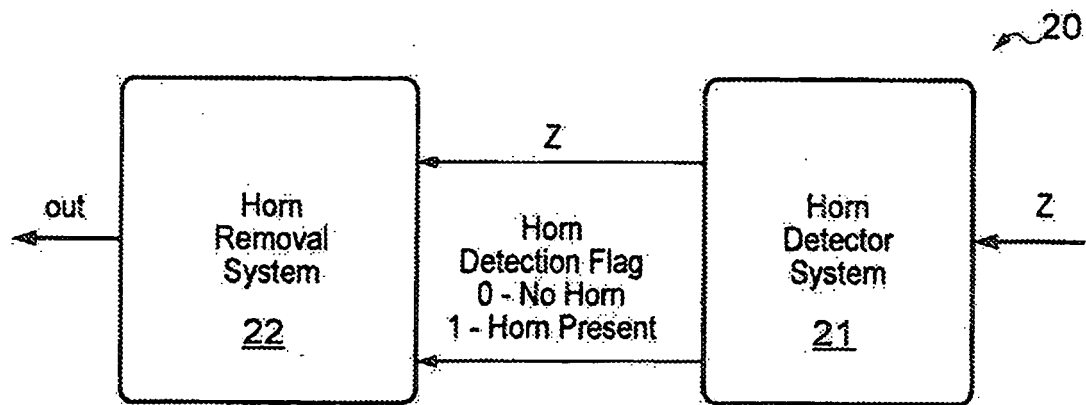


FIG. 2

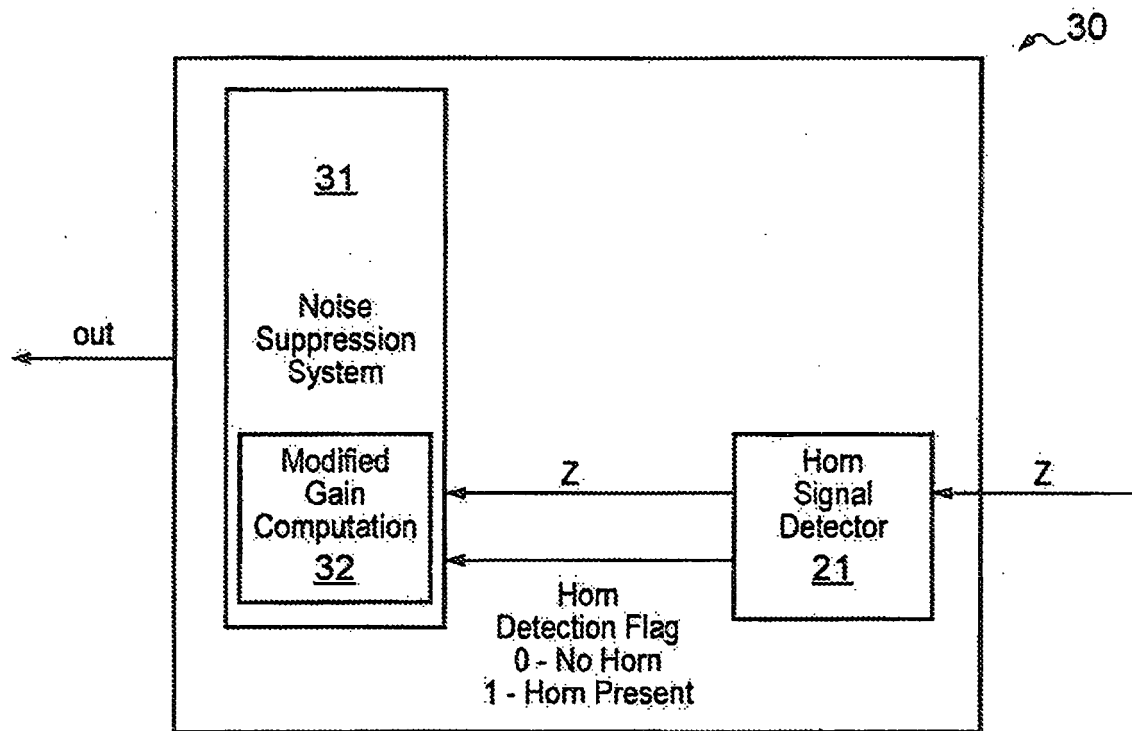


FIG. 3

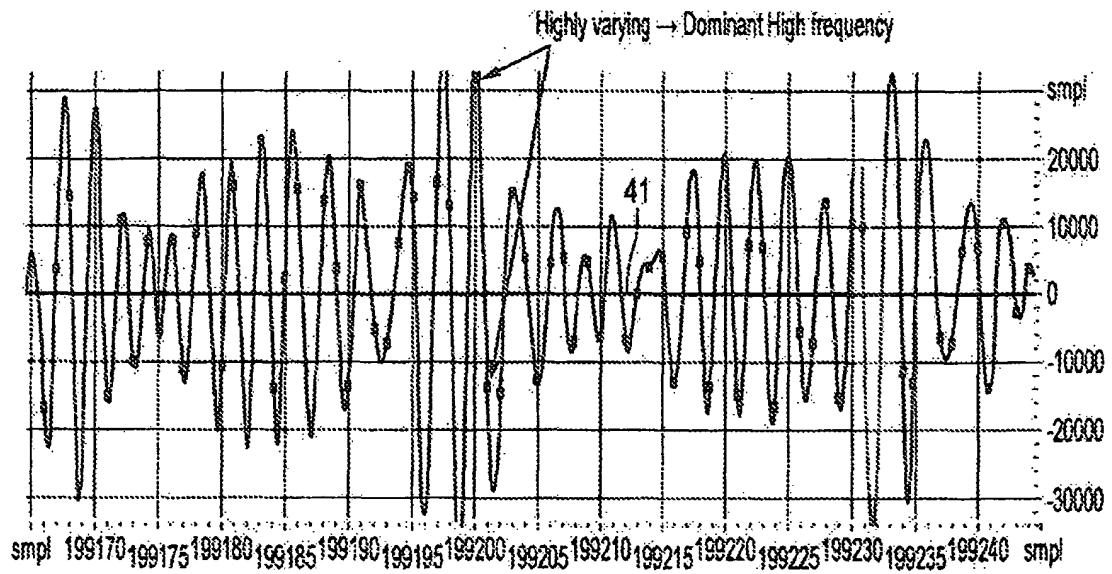


FIG. 4a

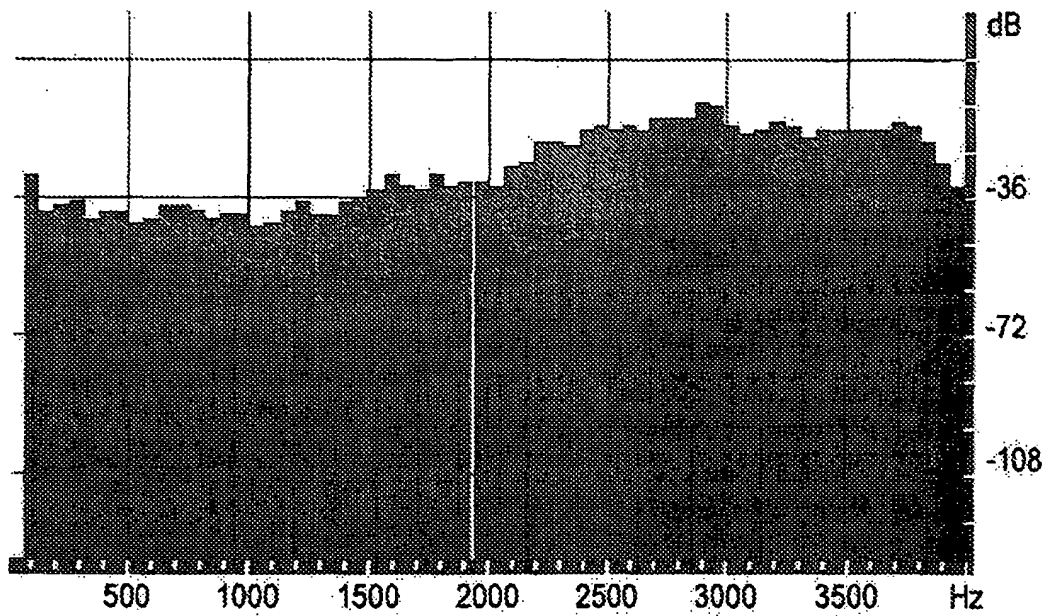


FIG. 4b

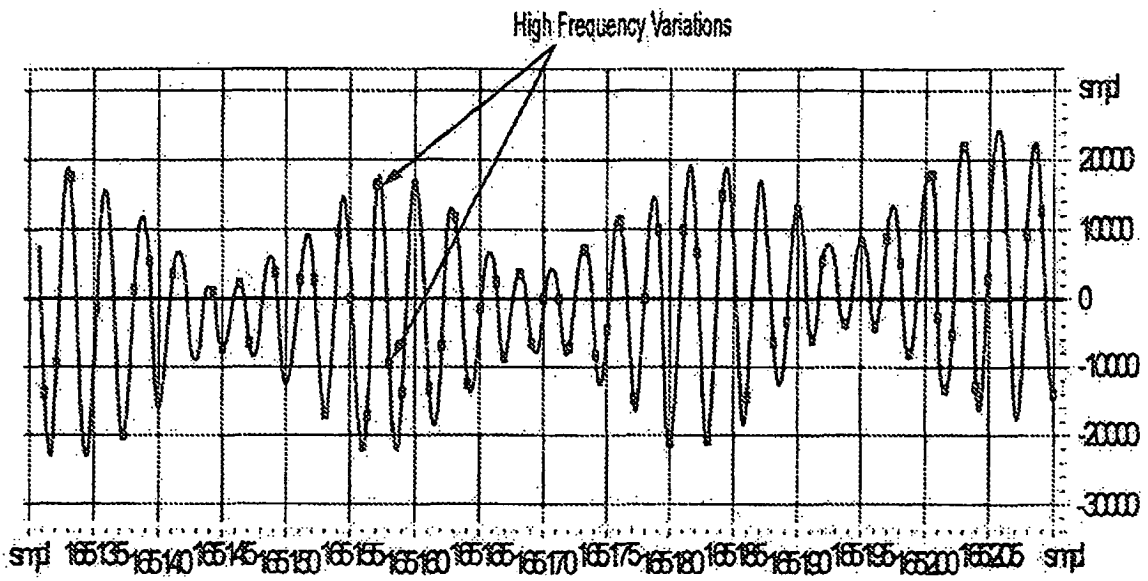


FIG. 5a

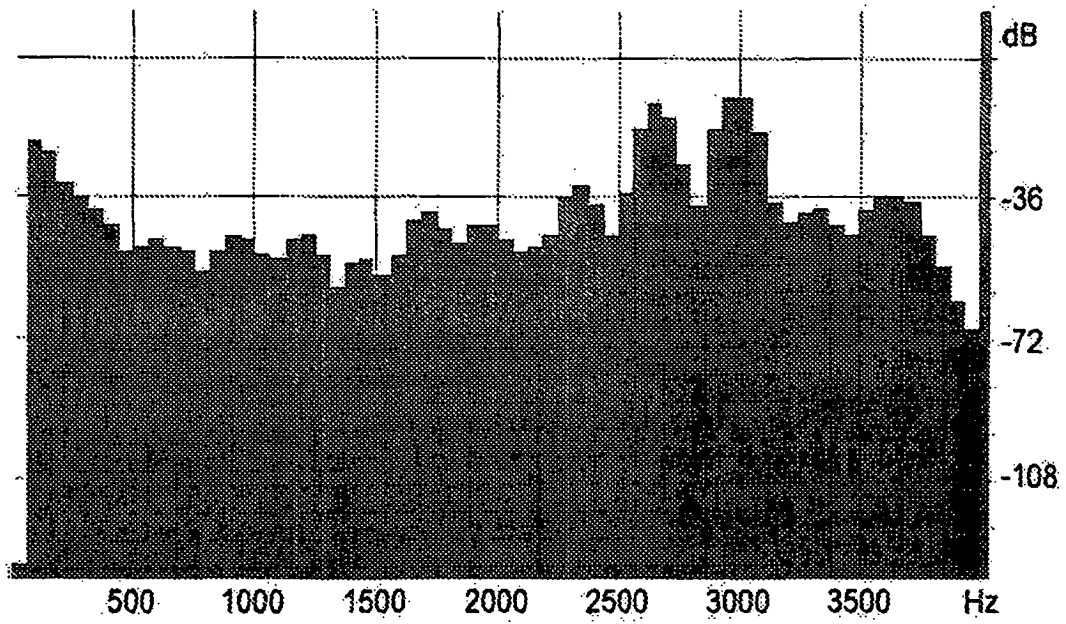


FIG. 5b

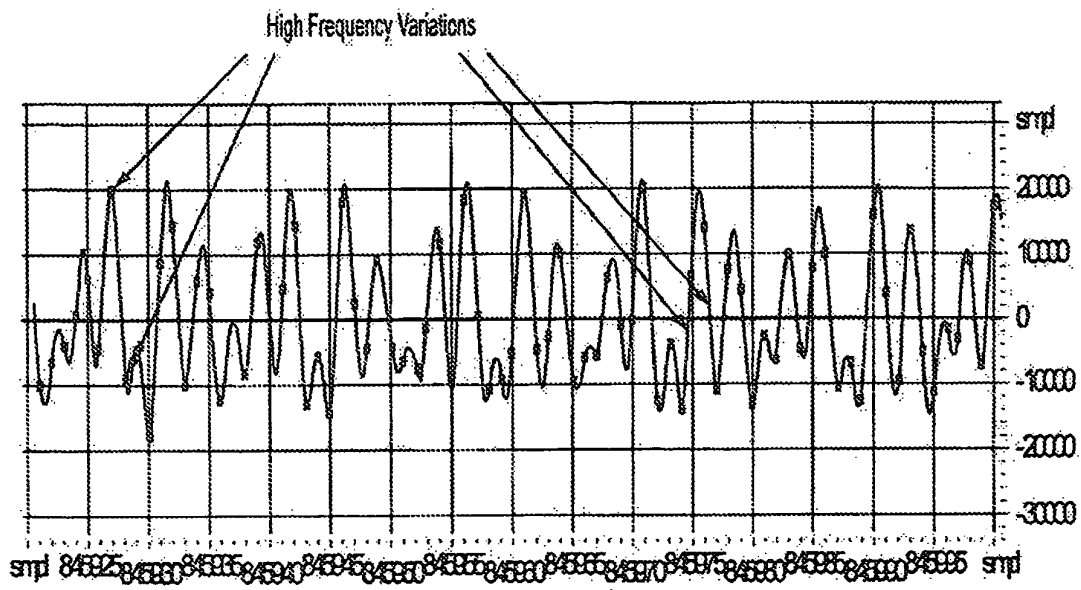


FIG. 6a

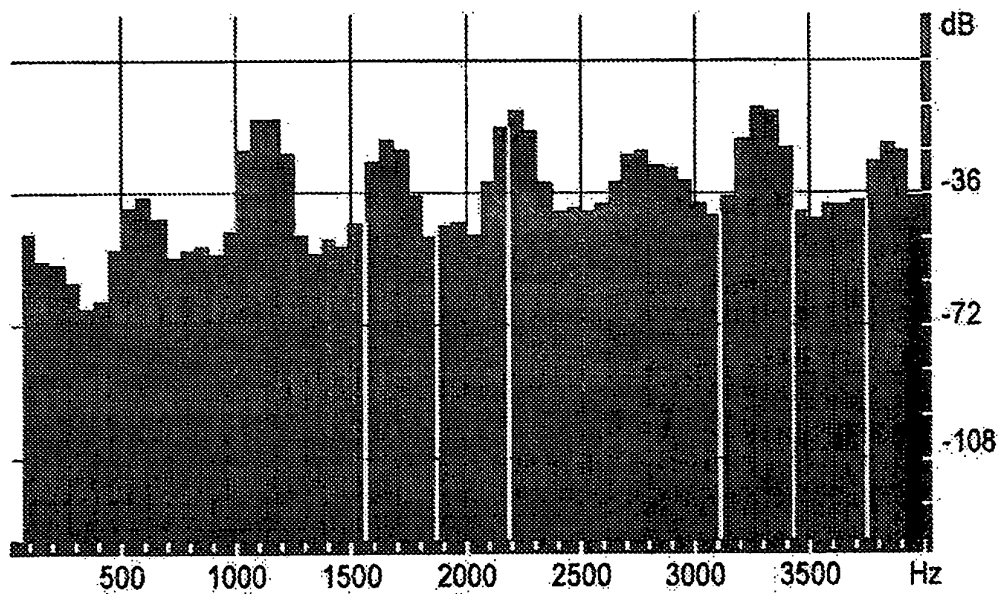


FIG. 6b

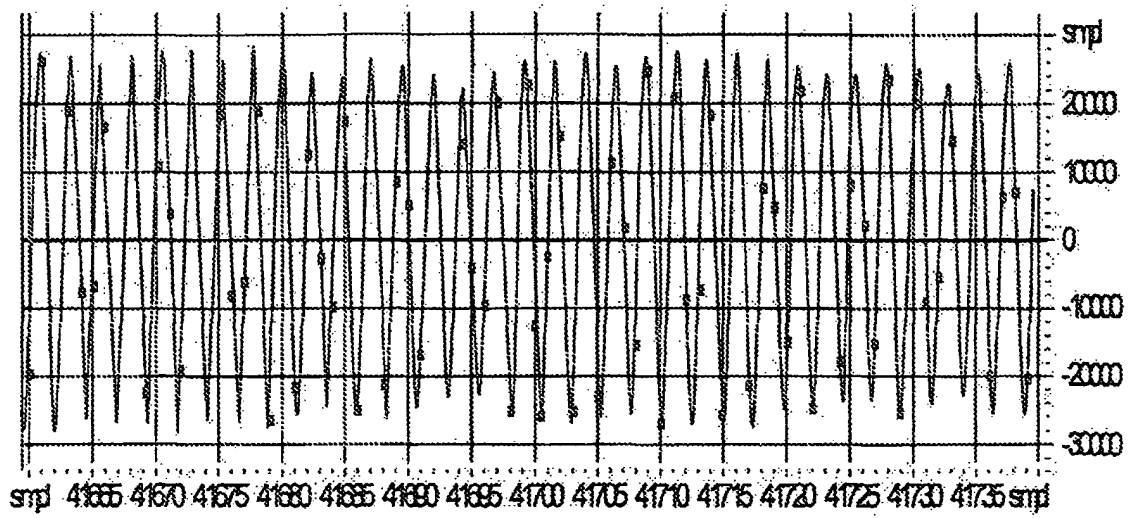


FIG. 7a

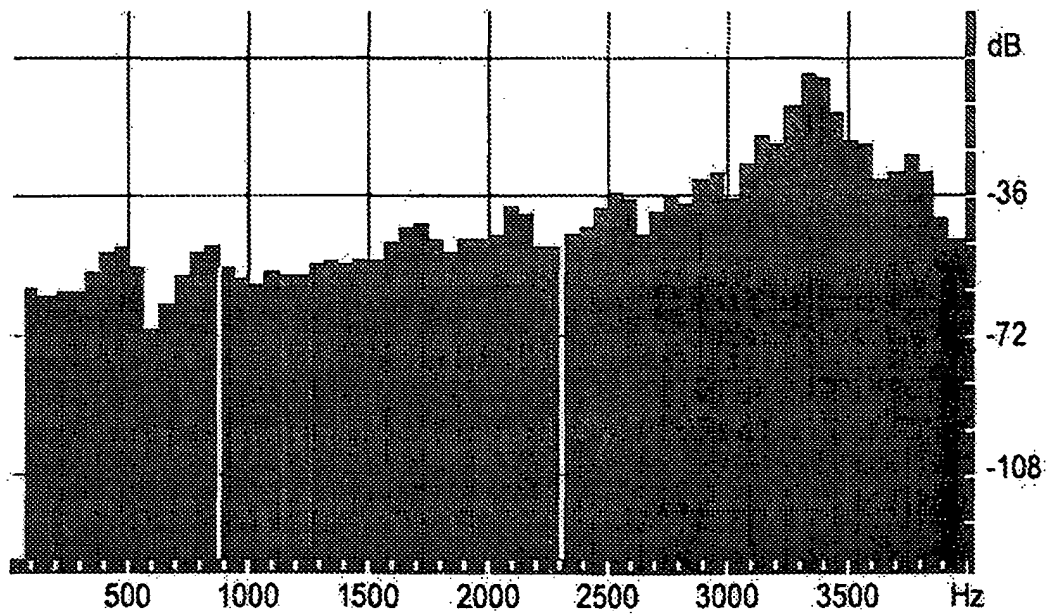


FIG. 7b

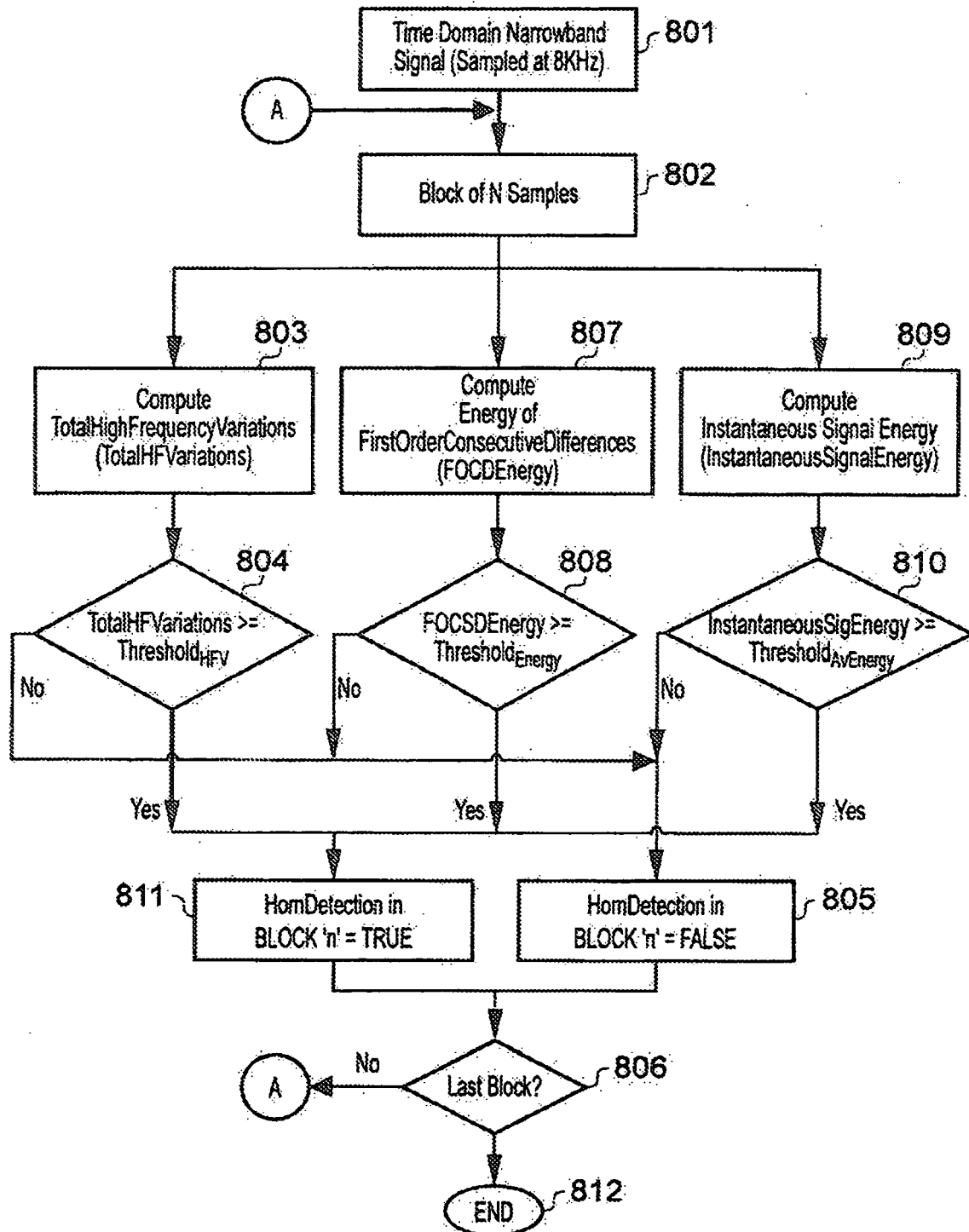


FIG. 8

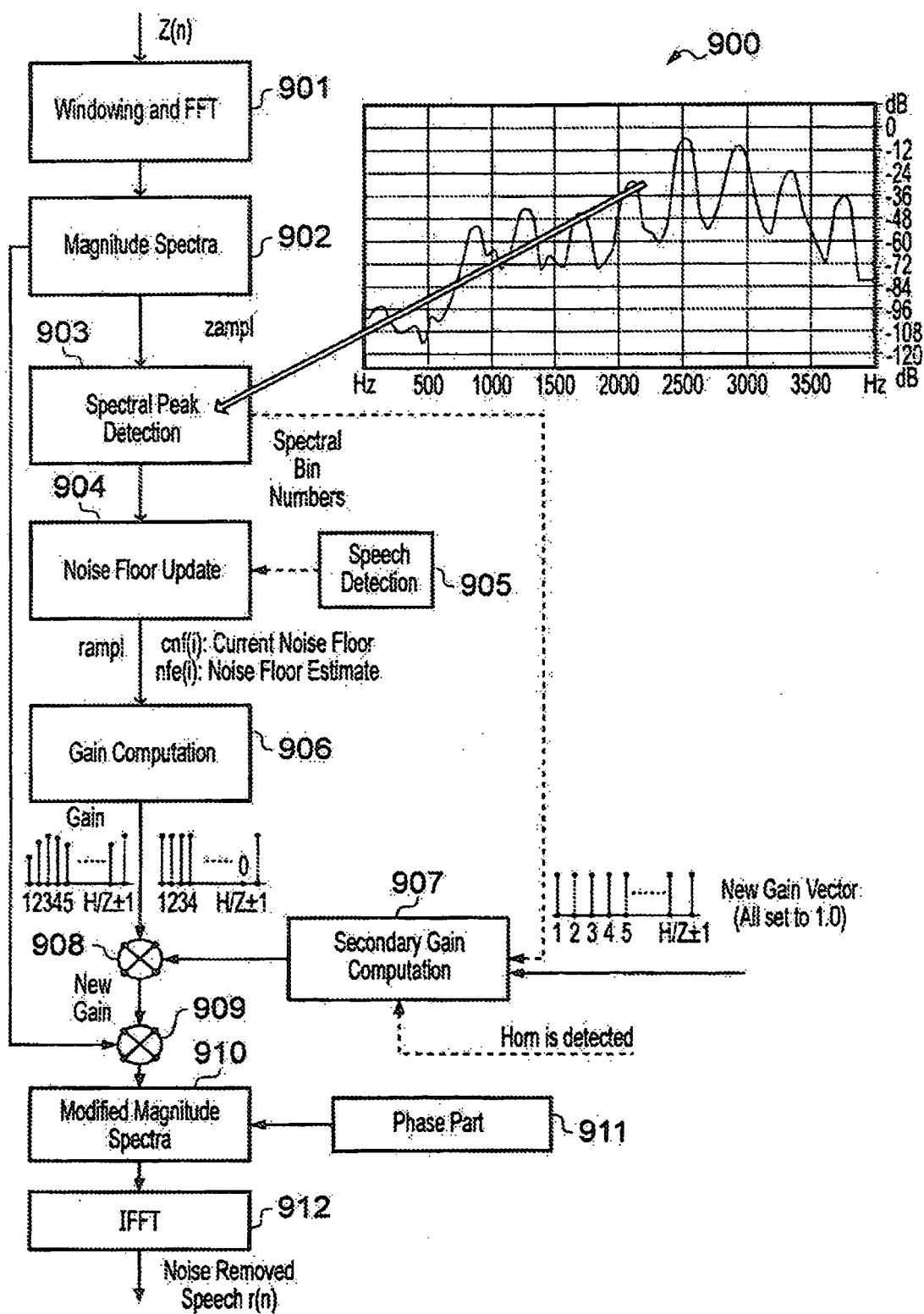


FIG. 9

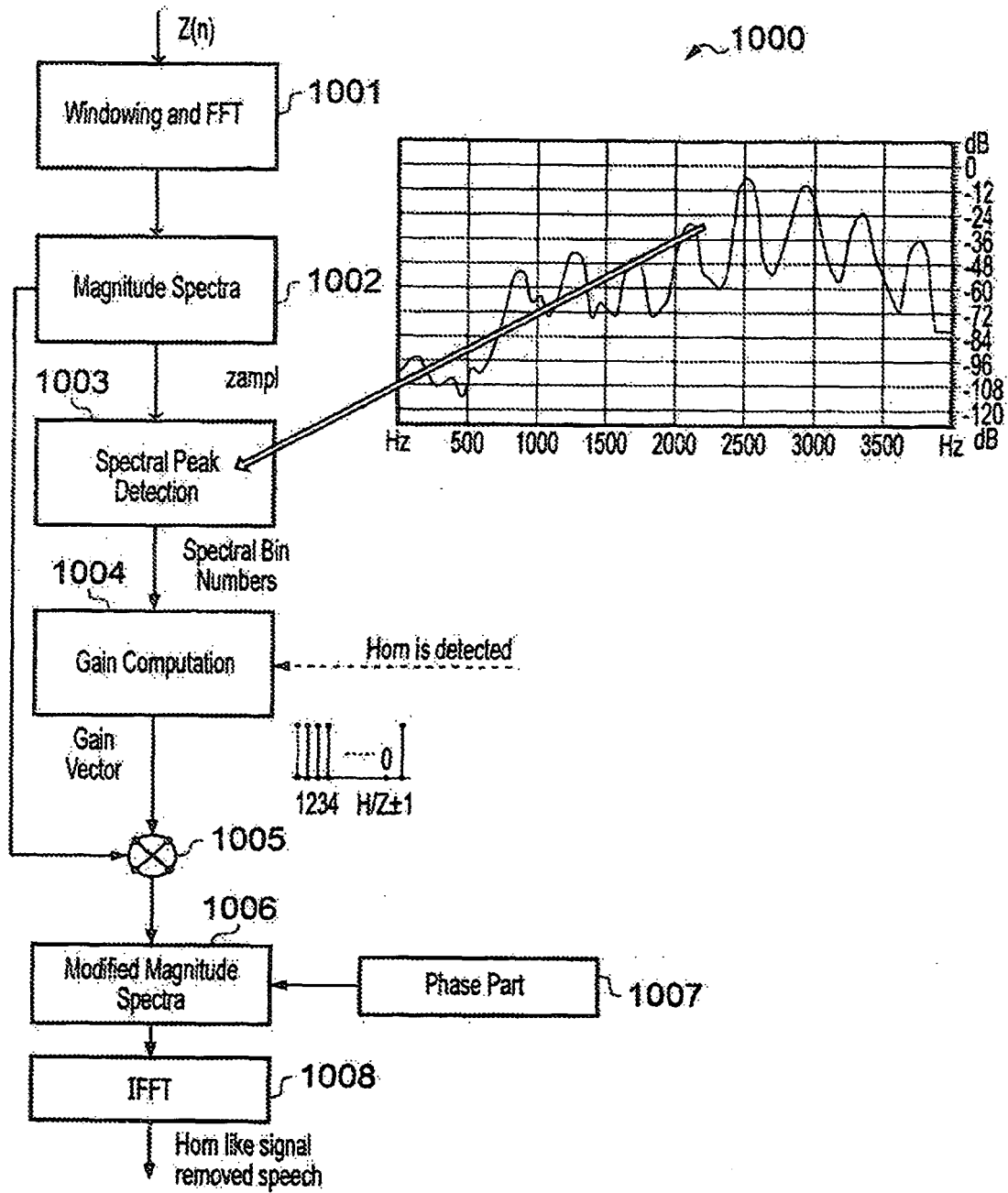


FIG. 10

**REFERENCES CITED IN THE DESCRIPTION**

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

**Patent documents cited in the description**

- US 20060200344 A [0013]
- US 20070192102 A [0014]

**Non-patent literature cited in the description**

- **R. MARTIN.** Spectral Subtraction Based on Minimum Statistics. *Signal Processing VII: Theories and Application*, 1994, 1182-1185 [0008]
- **M. SZCZERBA ; A. CZYZEWSKI.** Pitch Detection Enhancement Employing Music Prediction. *Journal of Intelligent Information Systems*, 2005, vol. 24 (2/3), 223-251 [0015]