



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
**27.05.2009 Bulletin 2009/22**

(51) Int Cl.:  
**G10L 21/02 (2006.01)**

(21) Application number: **07405332.3**

(22) Date of filing: **26.11.2007**

(84) Designated Contracting States:  
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC MT NL PL PT RO SE SI SK TR**  
Designated Extension States:  
**AL BA HR MK RS**

(72) Inventors:  
• **Dubuis, Baptiste**  
1018 Lausanne (CH)  
• **Zoia, Giorgio**  
1007 Lausanne (CH)

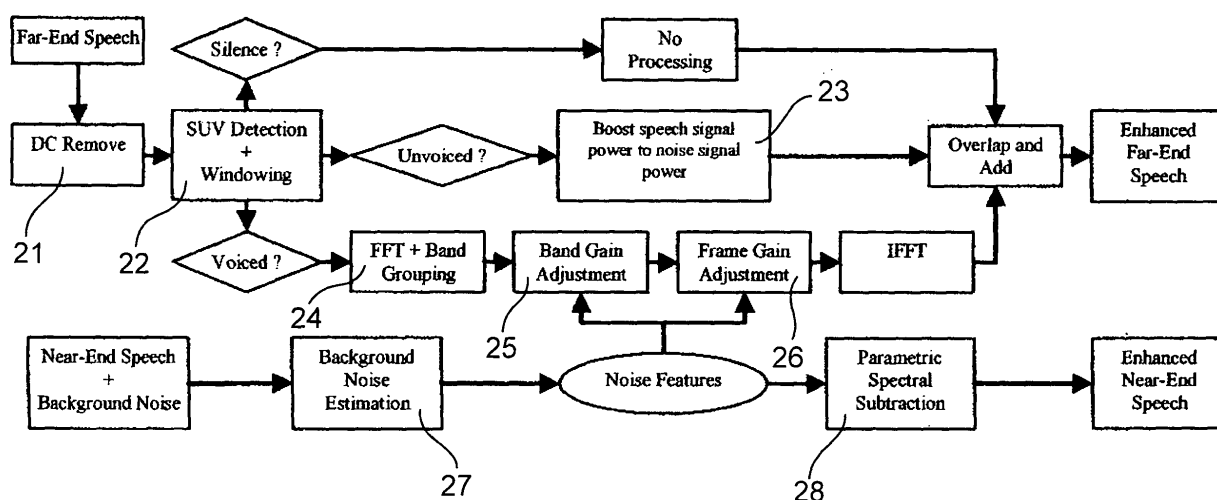
(71) Applicant: **EyeP Media S.A.**  
1020 Renens (CH)

(74) Representative: **Nithardt, Roland**  
**Cabinet Roland Nithardt**  
**Conseils en Propriété Industrielle S.A.**  
Y-Parc / Rue Galilée 9  
1400 Yverdon-les-Bains (CH)

(54) **Method and assembly to enhance the intelligibility of speech**

(57) The present invention concerns a method and an assembly designed to enhance the intelligibility of speech produced by a sound device in a noisy environment. The assembly comprises a microphone, or a telecommunication unit, which provides to a data processing module the voice of a remote speech. The data process-

ing module is designed to combine specific algorithms offering a perceptual improvement of the produced speech by increasing intelligibility, by saving an adequate signal quality and by saving as far as possible the overall power consumption. The enhanced speech as produced by the data processing module is then played in a speaker.



**FIG. 1**

**Description****Technical field**

5     **[0001]** The present invention concerns a method to enhance the intelligibility of speech produced by a sound device in a noisy environment.

**[0002]** The present invention also concerns an assembly for implementing this method to enhance the intelligibility of speech produced by a sound device in a noisy environment.

10    **Background Art**

**[0003]** Over the last decade, communication devices market has experienced a spectacular growth in terms of research, technology and users attention, especially the mobile or portable devices such as mobile phones, personal digital assistants or hearing aids.

15    **[0004]** The need to solve the problems of noise control and speech quality when dealing with small, low power device is critical.

**[0005]** In the history of communication, noise, in particular stationary background noise, has always been a problem. Every signal traveling from one point to another is prone to be corrupted by noise. Noise can come in various manners: from surrounding acoustic sources, such as traffic, babbling, reverberation or acoustic echo paths, or from electric/ electronic sources such as thermal noise. Background noise, also known as environmental noise, can seriously affect speech perceptual aspects such as quality or intelligibility. Therefore huge efforts have been produced during the last decades to overcome this problem.

**[0006]** A solution to the speech enhancement in the presence of local background noise is fundamental to the user experience. This issue is compounded by the consequences of possible usage in unfavorable environments and of rapid change in background conditions. Rapid means that those conditions may vary one or several times during the time of a normal conversation, even if this is a rather slow change in comparison to signal and noise frequencies so that noise can be mainly approximated as stationary in comparison. Automatic adaptation of perceptual aspects such as quality and especially intelligibility are then of the uppermost importance to provide as seamless as possible conversation and device use.

30    **[0007]** A classic noise reduction problem consists of reducing the level of stationary noise superimposed to a local voice (or sound in general) signal that is captured by the same recording device in the same time interval. On the other hand, remote voice signal arrives to a sound device more or less disturbed by remote background noise and local device noise, but it is added to local background noise only during the acoustic path from the device speaker to one ear and further disturbed by local background noise possibly reaching the other ear. This kind of noise cannot be reduced for the local user by signal processing in the digital domain; this can be obtained using the classic scheme only for the remote user. So, the only possible solution is to enhance the remote voice signal locally, in order to improve its perception when immersed in the local noisy condition.

**[0008]** If classic noise reduction constitutes a well-known branch of research and signal processing tools are mature enough to face it consistently in many cases, far-end speech enhancement in noisy condition is instead a relatively new issue. It is also trickier as it presents the necessity to compare signals and surrounding noise that cannot be captured by the very same device due to a dual-channel problem, and therefore are not so easy to compare in an objective manner.

40    **[0009]** One of the possible solutions concerns the change of volume, which is in fact not usable in any situation and in any place. Another solution is to use isolating headset devices. This solution is invasive and cannot be used everywhere. A conventional solution consists of changing location but it reduces the mobility and is not applicable in any case. A further solution consists of using noise canceling headsets. The drawback of such a solution is that it is invasive, needs extra battery and is costly.

**Disclosure of the Invention**

50    **[0010]** To overcome the above drawbacks of the prior art, an object of the present invention is to provide a method such as defined in preamble and characterized by a combination of specific algorithms offering a perceptual improvement of the produced speech by increasing intelligibility, by saving an adequate signal quality and by saving as far as possible the overall power consumption.

**[0011]** The method primarily adapts to non-personally invasive devices but it also operates on invasive devices.

55    **[0012]** The method applies especially when no direct or indirect control is possible on the source of background noise. It applies when the microphones of the device capture the background noise but not necessarily the source of speech, which may be local as well as remote, received through a communication link and rendered through the device speaker(s).

**[0013]** The field of use especially includes telecommunication devices, hearing aids devices and multimedia devices.

**[0014]** According to a preferred form of realisation, at least one algorithm is used for identifying signal segments as silence, voiced or unvoiced segments (SUV).

**[0015]** The unvoiced segments are processed by applying a constant amplification, given the reduced bandwidth of the voice signal and the corresponding high bandwidth of these unvoiced segments.

**[0016]** Advantageously, the silence segments are simply ignored.

**[0017]** According to an attractive form of the present invention, a band energy adaptation is especially conceived to avoid increases in the overall power of the long voiced segments. To this purpose, the overall power is redistributed where noise is less masking, with consequent reduction in the energy, instead of increasing it where noise is more intense.

**[0018]** Preferably, a certain amount of signal distortion is accepted to permit as advantage an increase in intelligibility in particular environmental conditions.

**[0019]** Specific approximations to theoretical algorithms are made in SUV segmentation, thresholds and band gain adjustments to reduce computation, allowing execution in real-time on portable devices and with consequent reduction in both CPU load and battery load of the sound device.

**[0020]** The object of the present invention is also achieved by an assembly for implementing this method as defined in the preamble and characterized in that said assembly comprises at least one microphone, one speaker, and a data processing module designed to combine specific algorithms offering a perceptual improvement of the produced speech by increasing intelligibility, by saving an adequate signal quality and by saving as far as possible the overall power consumption.

**[0021]** Advantageously, the data processing module comprises means designed to identify signal segments as silence, voiced and unvoiced segments. Preferably, this means is at least one algorithm.

**[0022]** For simplifying processing of unvoiced segments, the data processing module also comprises means designed to apply a constant amplification to said unvoiced segments, given the reduced bandwidth of the voice signal.

**[0023]** Furthermore the data processing module of the assembly may also comprise means designed to ignore the silence segments, and means designed to provide a band energy adaptation especially conceived to avoid increases in the overall power of the long voiced segment.

**[0024]** In a preferred embodiment of the assembly, the data processing module may comprise means designed to redistribute the overall power where noise is less masking instead of increasing it where noise is more intense, with consequent reduction in the energy consumed.

**[0025]** In order to reduce computation, with consequent reduction in both CPU load and battery load of the sound device, the assembly according to the present invention may comprise means designed to make specific approximations in SUV segmentation, thresholds and band gain adjustments.

### Brief Description of the Drawings

**[0026]** The present invention and its advantages will best appear in the following description of a mode of embodiment given as a non-limiting example and referring to the appended drawings, in which:

Figure 1 represents a block diagram for the overall speech enhancement method according to the present invention,

Figure 2 represents a block diagram for the SUV decision algorithm according to the method of the present invention,

Figure 3 represents a Bark filter bank usable for both noise and speech analysis according to the method of the present invention, and

Figure 4 represents a block diagram for the assembly according to the present invention.

### Best Mode for Carrying Out the Invention

**[0027]** The following subsections of the block diagram for the overall speech enhancement method according to the present invention such as illustrated by the Figure 1 will give a behavioral description of the different processing blocks, whereas the next section will describe in detail the implementation of each block. The noise estimation part is described in less detail as it constitutes a better known algorithm and is not relevant to the actual novelty of the proposal.

#### DC remove block 21

**[0028]** Voice signals captured through a microphone may contain a DC (continuous) component. Since signal processing modules are often based on energy estimation, it is important to remove this DC component in order to avoid useless very high offsets, especially in a limited numerical format case (16-bit integer). The DC remove filter implements a simple

IIR filter allowing the removal of the DC component inside the telephone narrow- and wide-band range limiting the loss in other low frequencies as far as possible.

#### SUV Detection block 22

**[0029]** A voice-only signal is typically composed by speech periods that are separated by Silence intervals. Moreover, speech periods can be subdivided into two classes, Unvoiced and Voiced sounds.

**[0030]** Speech periods are those when the talker is active. Roughly speaking, a speech sound can be considered as voiced if it is produced by the vibration of the vocal cords. Vowels are voiced sounds by definition. When a sound is instead pronounced so that it does not require the vocal cords to vibrate, it is called unvoiced. Only consonants can be unvoiced, but not all of them are. Silence normally refers to a period in the signal of interest where the talker is not speaking. But while not containing speech, most of the time the signal corresponding to "silence" regions rather different from zero as it can contain many kinds of interfering signals, such as background noise, reverberation or echo.

**[0031]** The SUV detection block 22 allows separating signal into silence, unvoiced and voiced periods. This is normally obtained by calculating a number of selected signal features, which are then weighted and fed to a suitable decision algorithm. As the whole algorithm works on a frame-by-frame basis, as often in signal processing for efficiency in computation, this block provides as output signal frames, each frame being windowed before processing (and frames are then overlapped at the end).

#### Speech Signal Simple Boost block 23

**[0032]** In terms of speech intelligibility, consonants, and therefore unvoiced sounds, often convey more important information than vowels do. Furthermore, unvoiced sounds are weaker than voiced sounds and are therefore more prone to be masked by noise.

**[0033]** Unvoiced signals nearly cover the entire speech band, which in most cases is approximately 3.5 or 7 kHz wide (8 or 16 kHz sampling rate). This allows boosting in a simple manner unvoiced portions by limiting at maximum the processing power. The enhancement is obtained by applying a gain in time domain to each sample so as to increase unvoiced speech power to a level at least equal to that of the background noise power. This has the effect of increasing the power of consonants against vowels.

#### Frequency Transform and Band Grouping block 24

**[0034]** The processing of the voiced part is the most expensive from a computation point of view: it requires analysis in the frequency domain. The frequency coefficients are preferably calculated by applying a Short-Time Fourier Transform (STFT) to the voiced speech signal. Once the coefficients computed, they are grouped into frequency bands to reflect in relative importance the nonlinear behavior of the human hearing. In fact, from a psychoacoustic point of view, critical bands increase in width as frequency increases. Grouping is obtained preferably using a Bark-like scale. The number of critical bands has been chosen to be preferably twenty-four, which trade-offs enough frequency resolution for the purpose of noise estimation, noise reduction and speech enhancement.

#### Band Gain Adjustment 25

**[0035]** After frequency transforming and grouping the signal into psycho-acoustically relevant critical bands (the same as done in noise analysis branch), the gain of each critical band is adjusted according to criteria that can result in an improvement of the overall intelligibility of voice periods of speech over noise. In particular, gain is increased inversely to the noise distribution in critical bands, which means signal is increased more where noise has less energy aiming at reinforcing SNR in bands that require a lower energy increase. Signal may even be reduced where noise is very strong to preserve as far as possible the energy level.

**[0036]** Improvement of intelligibility is often detrimental to speech quality (perceived quality in absence of background noise). To preserve good quality a number of thresholds are used to avoid:

- too much signal distortion when the signal-to-noise ratio is low, and
- too much useless distortion when noise is overall low,
- too much distortion after repartition of energy among critical bands.

These thresholds aim at preserving main timbre features, so that recognition of speaker is not compromised.

Frame Gain Normalization block 26

**[0037]** After the application of gains to each critical band of a signal frame, the frame gains are normalized depending on the power of the noise frame. If the original power of the speech frame was greater or equal than the power of the noise frame, then the energy of the signal is kept unchanged. But if the power of the noise frame was greater, then masking may occur. The speech frame power is boosted so that it has the same power as noise, taking care not to hit too high values leading to signal saturation.

**[0038]** After this normalization, signal is transformed back to the time domain and overlap-and-add is applied to frames to recreate a complete signal (with silence, unvoiced and voiced parts all together again).

Background Noise Estimation and Features Extraction block 27

**[0039]** Background Noise Estimation consists of separating to background noise captured locally by the device microphone from noise + speech periods. Many algorithms exist for this kind of separation. A voice activity detection (VAD) is preferably used here to separate pure noise segments and the noise features are extracted as explained above by frequency transform and grouping into critical bands. Noise energy for each critical band is used by the enhancement algorithm outlined above.

Parametric Spectral Subtraction block 28

**[0040]** Parametric Spectral Subtraction is the core of the noise reduction algorithm that can be applied to the local speech signal before transmission to the remote peer. This part has no influence on the remote speech enhancement. In any case, gains are calculated according to an Ephraim-Malah algorithm.

**[0041]** The proposed application preferably targets mobile device implementations. As such, important limitations are imposed by the device and CPU in comparison to theoretical solutions and many approximations may be necessary to reduce the computational complexity while saving the result accuracy.

**[0042]** The following paragraphs describe examples of approximations which are preferably made in SUV segmentation, thresholds and band gain adjustments to reduce computation, with consequent reduction in both CPU load and battery load.

Fixed-point proposed implementation example

**[0043]** The proposed implementation example runs completely in fixed-point arithmetic. Signals are signed short integers (16-bit dynamic range), whereas internal coefficients for frequency transforms and other analyses are 32-bit fixed-point numbers. Precision of fixed-point numbers will be detailed later in this document when important.

**[0044]** In terms of numerical operations, solutions are proposed too in order to avoid division and modulo operators at least on a sample-by-sample basis, since these functions are often not available in device instruction sets and are consequently realized in software using hundreds or thousands of CPU cycles.

**[0045]** The following paragraphs replicate the structure of the overall process description and contain detail about the specific fixed-point arithmetic algorithm implementation and specific filter and formula aspects.

**[0046]** The DC Remove filter block 21 is applied to the audio signal frames before processing. In order to save CPU resources, and since microphone characteristics are often poor at low frequencies in mobile devices, a simple high-pass, fixed-point IIR filter is used. Cutoff frequency is approximately 200 Hz in narrowband, 60 Hz in wider bands.

SUV segmentation

**[0047]** To segment the audio signal into a silence, unvoiced or voiced portions, three different features are considered, the log-energy, the normalized autocorrelation coefficient and the zero-crossing count.

The log-energy is computed as:

**[0048]**

$$E_s = 10 \times \log_{10} \left( \varepsilon + \frac{1}{N} \sum_{n=1}^N s^2(n) \right)$$

where  $s(n)$  is the signal sample,  $N$  is the number of samples per frame (20 ms frame for example) and  $\varepsilon$  is a small constant to avoid log of 0. After log calculation log-energy values may be stored in signed 7b/8b (16-bit) numbers.

**[0049]** The normalized autocorrelation coefficient at unit sample delay is approximated as:

$$C_1 = \frac{\sum_{n=1}^N s(n)s(n-1)}{\sum_{n=1}^N s^2(n)}$$

Voiced sounds are more concentrated at low frequencies, and then normalized autocorrelation tends to be higher (near to 1) for voiced than unvoiced segments. The denominator sum is an approximation of the correct formula to avoid calculation of the square root. The range is of course -1 to 1 (signed 0b/15b representation).

**[0050]** The number of zero-crossings for a frame is computed as:

$$N_z = \sum_{n=1}^N |\text{sgn}(s(n)) - \text{sgn}(s(n-1))|$$

where  $\text{sgn}$  is the sign operator. The number of zero-crossing is an integer value (15b/0b) representation.

**[0051]** Figure 2 represents the block diagram for the SUV decision algorithm. To decide to which class (S, U or V) the segment belongs, a distance is computed between the actual feature vector and each of the three classes. This is done by assuming that the features for each class belong to a multidirectional Gaussian distribution with known mean vector and covariance matrices  $W_i$ , corresponding respectively to the class voiced, unvoiced and silence. The index  $i$  is 1, 2 or 3 for the three classes.

**[0052]** Mean vectors and covariance matrices for the three classes are obtained (trained) by a given database of speech utterances. The data is segmented manually into silence, voiced and unvoiced, and then for each of these segments the three features above are calculated.

**[0053]** Once mean vectors and covariance matrices are available, the decision is taken according to the scheme of the Figure 2, where  $d_i$  is the error to be minimized in classical minimum probability-of-error decision rule, and then:

$$d_i = (\vec{x}_i - \vec{m}_i)' \cdot W_i^{-1} \cdot (\vec{x}_i - \vec{m}_i)$$

being  $x$  the feature vector,  $m$  the mean vector and  $W$  the covariance matrix.

**[0054]** Instead of using all features with the same weight to discriminate among classes, the following procedure is used as shown in the block diagram. First the segment is tested for Voice class using the log-energy and the zero-crossing count. If the resulting distance  $d_1$  is minimal among the three distances, and if the log-energy is higher than a given threshold, then Voice is decided. If the log-energy is lower than the threshold, then Silence is decided. The threshold has to be determined empirically. The actual value of the threshold is preferably 3'900, relative to the 7b/8b format described above for log-energy precision.

**[0055]** If  $d_1$  is not minimal, then the distance  $d_3$  from the silence class with the autocorrelation feature only is calculated. If it is minimal, then Silence is decided, otherwise Unvoiced is decided.

#### Speech Signal Simple Boost

**[0056]** Calling the power of the speech signal  $P_s$ , the power of the noise signal  $P_w$ , the signal-to-noise ratio  $\text{SNR} = P_s/P_w$ , the enhancement of unvoiced segments is simply obtained applying a gain in time domain to each sample to increase the signal power to a level at least equal to that of the noise power.

**[0057]** The simple boost can be described for each sample as follows:

$$s_{enhanced}(n) = \begin{cases} s(n), SNR \geq 1 \\ \min(T_{unvoiced}, \frac{1}{\sqrt{SNR}}) \cdot s(n), SNR < 1 \end{cases}$$

**[0058]** The parameter  $T_{unvoiced}$  is an adaptive threshold that avoids saturation. For each frame, the threshold is calculated as the maximum given by the chosen representation (32-bit) over the actual frame energy.

**[0059]** After the STFT, frequencies are grouped into frequency bands (according to human hearing) using a Bark-like scale as represented by the Figure 3. The following formula is used for the single frequencies:

$$Bark = 13.1 \cdot \arctan(0.00074 \cdot f) + 2.24 \cdot \arctan(1.85 \cdot 10^{-8} \cdot f^2) + 10^{-4} \cdot f$$

**[0060]** The number of band-pass filters, and therefore the number of critical band is twenty-four, the result as shown in Figure 3.

#### Band Gain Adjustment

**[0061]** Signal-to-Noise ratio in the frequency domain is defined as:

$$SNR = \frac{\sum_{m=1}^M |S(m)|^2}{\sum_{m=1}^M |W(m)|^2} = \frac{P_s}{P_w}$$

where  $S$  and  $W$  are the STFTs of signal and noise respectively. To avoid useless calculation, power in the frequency domain is simply taken from the one in the time domain by the following well-known theorem:

$$\sum_{n=1}^N |s(n)|^2 = \frac{1}{N} \sum_{k=1}^N |S(k)|^2$$

**[0062]** Furthermore, given the twenty-four critical bands  $B_i$ , the Noise Repartition Ratio for the  $i^{\text{th}}$  band is calculated by the following formula:

$$NRR_i = \frac{\sum_{b \in B_i} |W(b)|^2}{\sum_{m=1}^M |W(m)|^2}$$

**[0063]** The adjustment gain for each speech band is calculated as follows:

$$G(i) = \alpha + \beta \cdot SNR + \gamma \cdot \min\left(\frac{1}{NRR_i}, T\right)$$

with the timbre variation bias  $\alpha$  that has a value of 0.5, the SNR reference factor  $\beta$  has a value of 3, the noise factor  $\gamma$  has a value of 12.

**[0064]** This last formula is in theory one of the most critical parts of the algorithm since the computation of the inverse of the NRR can be very costly as it would require one integer division per critical bands. This has some consequences in mobile devices. Therefore a different solution is used in practice than the flat division. A property of logarithms is used:

$$\log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y)$$

so that:

$$\frac{1}{x} = 2^{\log_2(1) - \log_2(x)}$$

**[0065]** The choice of base 2 is made for efficiency reasons with simple instruction sets (such as those of portable devices). In fact, the exponential can be obtained by a left shift of the necessary positions (since the binary format is used), whereas the  $\log_2$  can be approximated by the following pseudo-code:

```

r=0;
  if (x>=65536)
  {
    x>>=16;
    r += 16;
  }
  if(x>=256)
  {
    x >>= 8;

    r+=8;
  }
  if (x>=16)
  {
    x >>= 4;
    r+=4;
  }
  if (x>=4)
  {
    x>>=2;
    r+=2;
  }
  if (x>=2)
  {
    r += 1;
  }
  Result = r;

```



[0066] The result is approximated but computation is reduced by a factor 15. Using this algorithm, the threshold T has an actual value of 256.

#### Frame Gain Normalization

[0067] Gains are normalized using the following equation:

$$G'(i) = \begin{cases} G(i), & P_s \geq P_w \\ \min\left(T_{voiced}, \sqrt{\frac{P_w}{P_s}} \cdot G(i)\right), & P_s < P_w \end{cases}$$

If the power of the noise frame was greater than signal originally, then masking is more likely to occur. It is then necessary to boost the speech frame power so that it has the same power as noise. A threshold  $T_{voiced}$  is set based on the initial power of the signal to avoid saturation. (the same as  $T_{unvoiced}$  is estimated above).

#### Background Noise Estimation and Features Extraction

[0068] Background noise is analyzed in the same way as the remote signal is, that is STFT is calculated and noise power is calculated for each critical band as explained above. The twenty-four noise coefficients are passed to the enhancement algorithm to proceed with SNR calculation and gain modifications for unvoiced and voiced segments.

[0069] Figure 4 represents the block diagram of the assembly 10 according to the present invention and shows how the different elements are connected. The source of the voice can be either a local microphone 11, or optionally a telecommunication unit 12, which provides to a data processing module 13 the voice of a remote speech. The data processing module 13 is used to combine specific algorithms offering a perceptual improvement of the produced speech by increasing intelligibility, by saving an adequate signal quality and by saving as far as possible the overall power consumption. The enhanced speech as produced by the data processing module 13 is played in a speaker 14. The telecommunication unit 12 has the capability to connect to a remote system that is a source of speech, especially a telecommunication device, and is optional.

#### Claims

1. Method to enhance the intelligibility of speech produced by a sound device in a noisy environment, **characterized by** a combination of specific algorithms offering a perceptual improvement of the produced speech by increasing intelligibility, by saving an adequate signal quality and by saving as far as possible the overall power consumption.
2. Method according to claim 1, **characterized in that** at least one algorithm is used for identifying signal segments as silence, voiced or unvoiced segments.
3. Method according to claim 2, **characterized in that** the processing of unvoiced segments is simplified by applying a constant amplification to said unvoiced segments, given the reduced bandwidth of the voice signal.
4. Method according to claim 2, **characterized in that** the silence segments are ignored.
5. Method according to claim 1, **characterized in that** a band energy adaptation is especially conceived to avoid increases in the overall power of the long voiced segment.
6. Method according to claim 5, **characterized in that** the overall power is redistributed where noise is less masking instead of increasing it where noise is more intense, with consequent reduction in the energy consumed.
7. Method according to claim 1, **characterized in that** a certain amount of distortion is accepted to permit an increase in intelligibility in particular environmental conditions.
8. Method according to claim 1, **characterized in that** specific approximations are made in SUV segmentation, thresholds and band gain adjustments to reduce computation, with consequent reduction in both CPU load and battery

load of the sound device .

- 5 9. Assembly to enhance the intelligibility of speech produced by a sound device in a noisy environment, this assembly being designed for implementing the method according to claims 1 to 8, **characterized in that** said assembly (10) comprises at least one microphone (11), one speaker (14), and a data processing module (13) designed to combine specific algorithms offering a perceptual improvement of the produced speech by increasing intelligibility, by saving an adequate signal quality and by saving as far as possible the overall power consumption.
- 10 10. Assembly according to claim 9, **characterized in that** the data processing module (13) comprises means designed to identify signal segments as silence, voiced and unvoiced segments.
11. Assembly according to claim 10, **characterized in that** the means designed to identify signal segments as silence, voiced and unvoiced segments is at least one algorithm.
- 15 12. Assembly according to claim 9, **characterized in that**, for simplifying the processing of the unvoiced segments, the data processing module (13) comprises means designed to apply a constant amplification to said unvoiced segments, given the reduced bandwidth of the voice signal.
- 20 13. Assembly according to claim 9, **characterized in that** the data processing module (13) comprises means designed to ignore the silence segments.
- 25 14. Assembly according to claim 9, **characterized in that** the data processing module (13) further comprises means designed to provide a band energy adaptation especially conceived to avoid increases in the overall power of the long voiced segment.
- 30 15. Assembly according to claim 14, **characterized in that** the data processing module (13) comprises means designed to redistribute the overall power where noise is less masking instead of increasing it where noise is more intense, with consequent reduction in the energy consumed.
- 35 16. Assembly according to claim 9, **characterized in that** the data processing module (13) comprises means designed to make specific approximations in SUV segmentation, thresholds and band gain adjustments to reduce computation, with consequent reduction in both CPU load and battery load of the sound device.
- 40
- 45
- 50
- 55

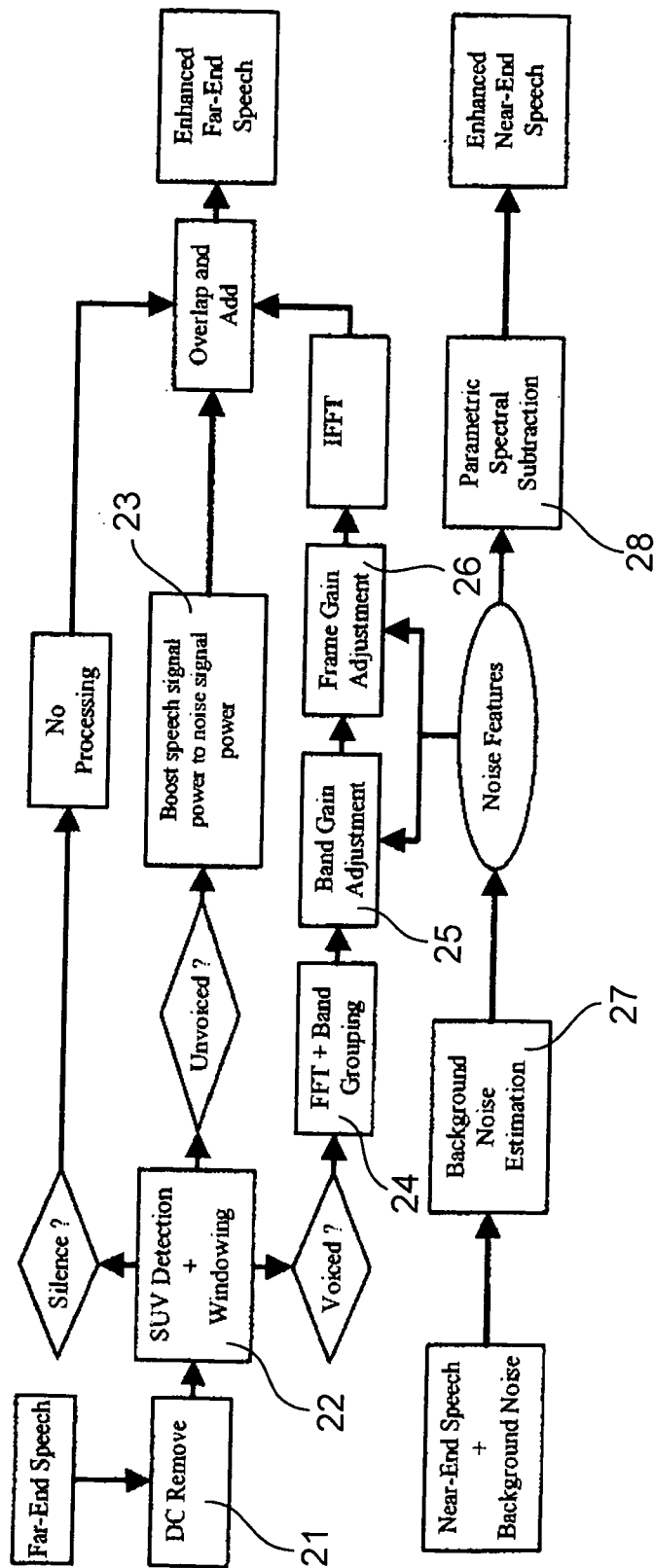


FIG. 1

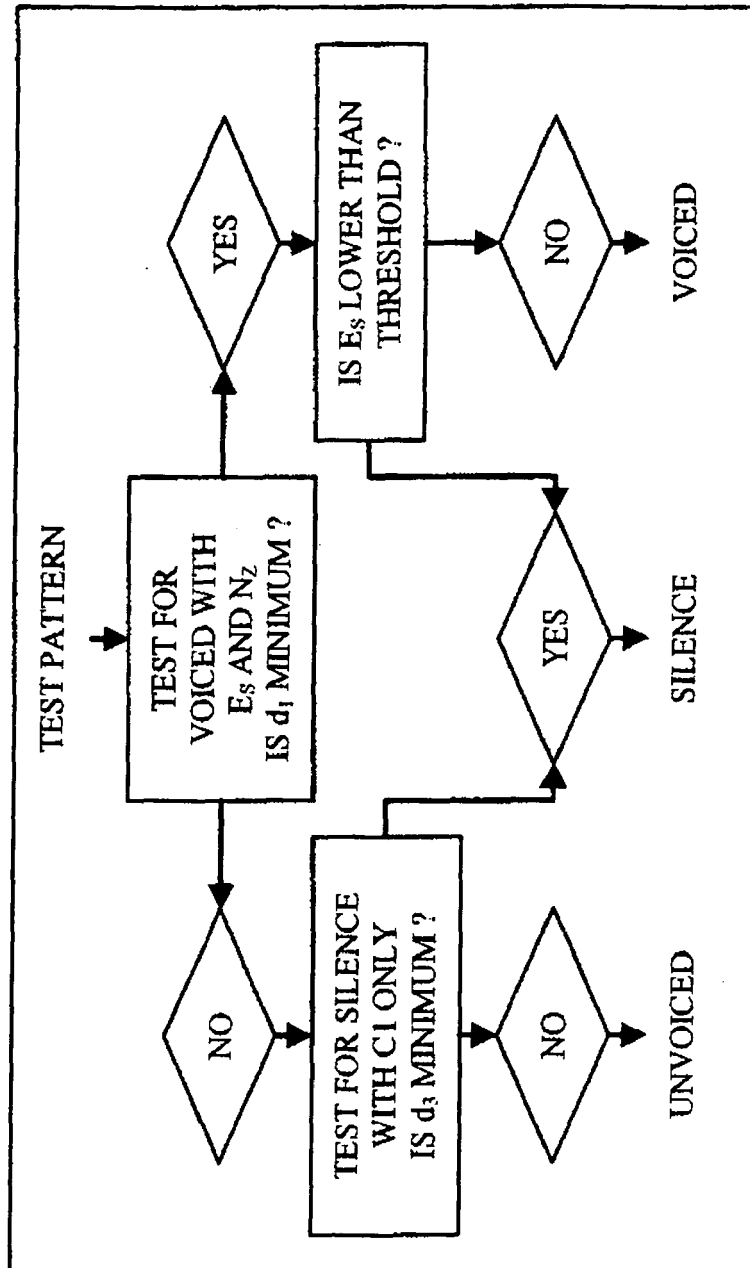


FIG. 2

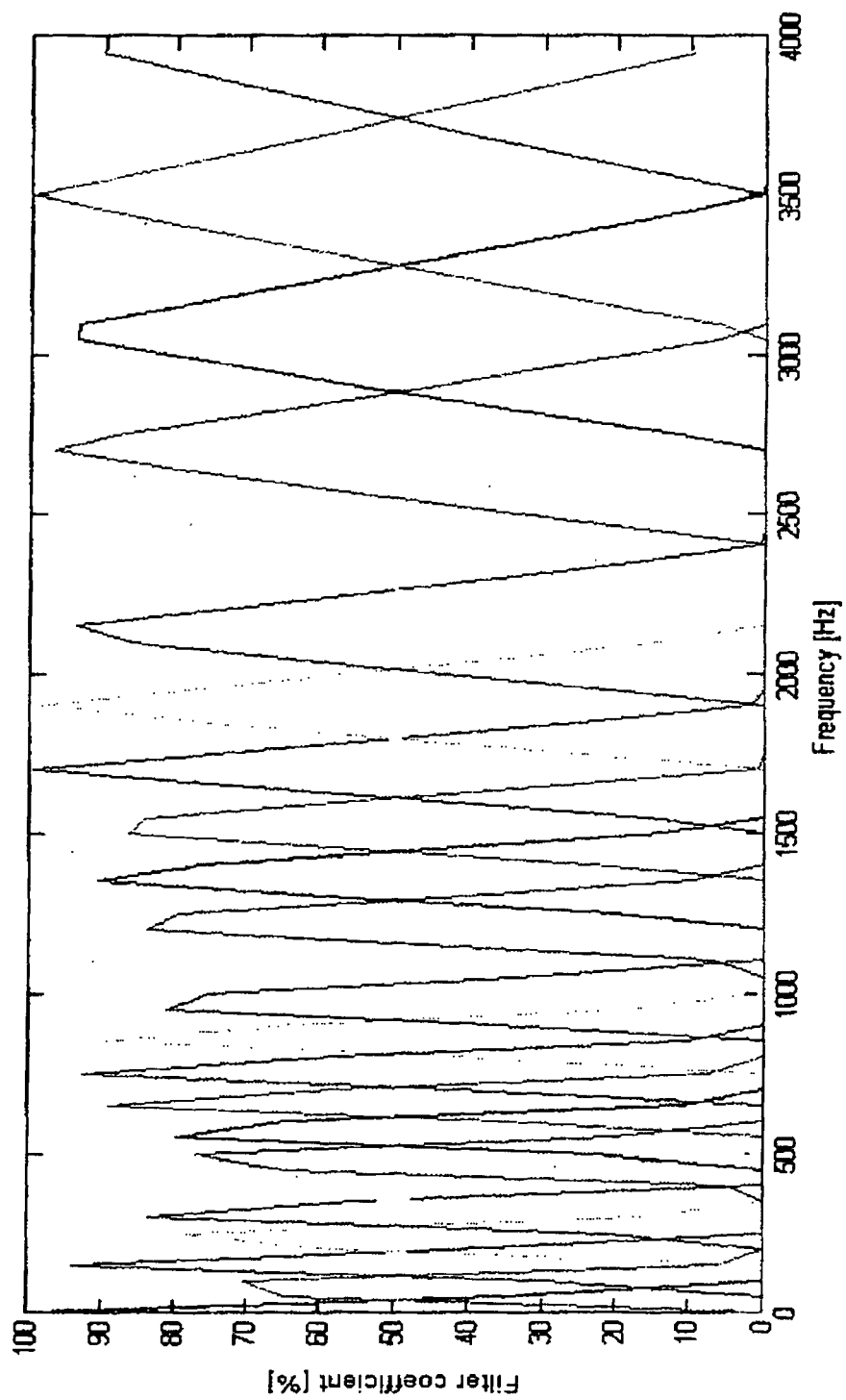


FIG. 3

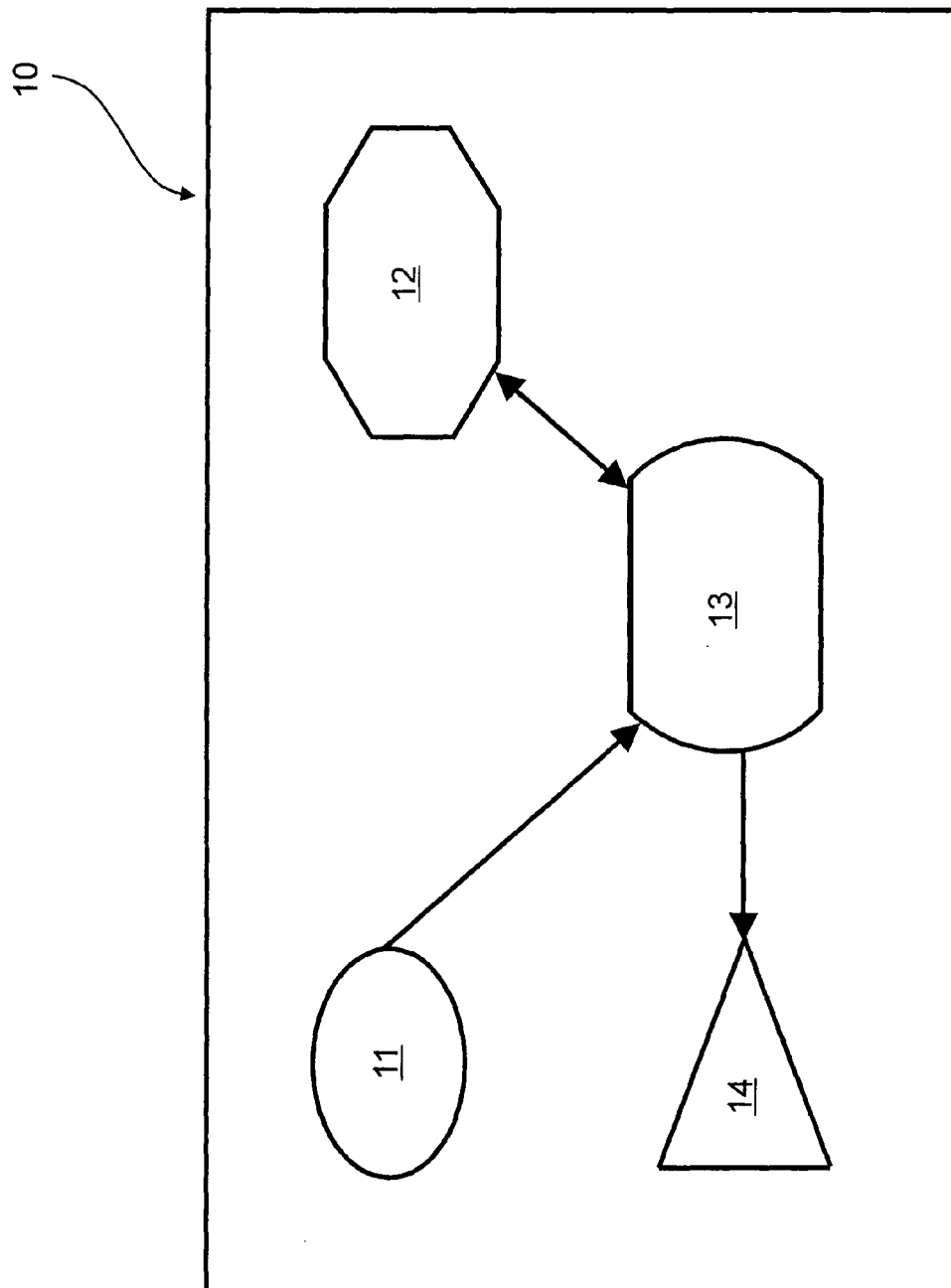


FIG. 4



DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	LEE S.H ET AL: "Real Time Speech Intelligibility enhancement based on the background noise analysis" PROCEEDINGS OF FOURTH IASTED "INTERNATIONAL CONFERENCE SIGNAL PROCESSING, PATTERN RECOGNITION AND APPLICATIONS", 14 February 2007 (2007-02-14), pages 287-292, XP002472964 INNSBRUCK, AUSTRIA * the whole document *	1-16	INV. G10L21/02
A	ATAL B S ET AL: "A PATTERN RECOGNITION APPROACH TO VOICED-UNVOICED-SILENCE CLASSIFICATION WITH APPLICATIONS TO SPEECH RECOGNITION" IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, IEEE INC. NEW YORK, US, vol. ASSP-24, no. 3, June 1976 (1976-06), pages 201-212, XP009040248 ISSN: 0096-3518 * abstract * * page 201, right-hand column, last paragraph - page 205, left-hand column, last paragraph *	1,2,8-11	TECHNICAL FIELDS SEARCHED (IPC) G10L
A	VIRAG N: "Speech enhancement based on masking properties of the auditory system" ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 1995. ICASSP-95., 1995 INTERNATIONAL CONFERENCE ON DETROIT, MI, USA 9-12 MAY 1995, NEW YORK, NY, USA, IEEE, US, vol. 1, 9 May 1995 (1995-05-09), pages 796-799, XP010625353 ISBN: 0-7803-2431-5 * abstract; figure 1 *	1,5-9, 14-16	
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 17 March 2008	Examiner Chétry, Nicolas
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			



European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 07 40 5332

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	<p>BEROUTI M ET AL: "ENHANCEMENT OF SPEECH CORRUPTED BY ACOUSTIC NOISE" INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH &amp; SIGNAL PROCESSING. ICASSP. WASHINGTON, APRIL 2 - 4, 1979, NEW YORK, IEEE, US, vol. CONF. 4, 1979, pages 208-211, XP001079151 * abstract; figure 1 * -----</p>	1,5,9, 14-16	
			TECHNICAL FIELDS SEARCHED (IPC)
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 17 March 2008	Examiner Chétry, Nicolas
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons ..... &amp; : member of the same patent family, corresponding document</p>			

2  
EPO FORM 1503 03.82 (P04001)