(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 521 243 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
06.04.2005 Bulletin 2005/14

(51) Int Cl.[7]: G10L 21/02, G10L 19/04

(21) Application number: 03022251.7

(22) Date of filing: 01.10.2003

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

(71) Applicant: SIEMENS AKTIENGESELLSCHAFT
80333 München (DE)

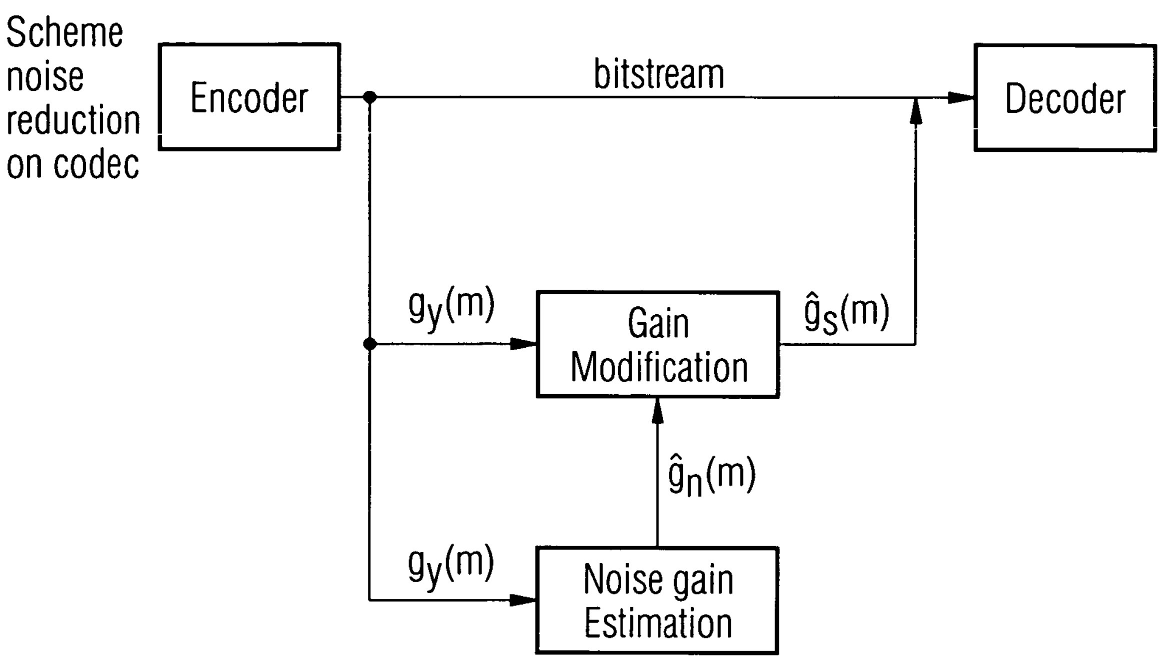
(72) Inventors:
- Beaugeant, Christophe, Dr.
81737 München (DE)
- Dütsch, Nicolas
80809 München (DE)
- Heiss, Herbert, Dr.
82178 Puchheim (DE)
- Taddei, Hervé, Dr.
81543 München (DE)

# (54) Speech coding method applying noise reduction by modifying the codebook gain

(57) The invention refers to a method for encoding an acoustic signal (y(n)) containing a speech component and a noise component by using an analysis through synthesis method, wherein for encoding the acoustic signal a synthesised signal is compared with the acoustic signal for a time interval, said synthesised signal being described by using a fixed codebook and an associated fixed gain, comprising the steps:

- Determining a fixed gain ($g_y(m)$) of the acoustic signal (y(n)) for the time interval;
- Extracting an estimated fixed gain ($\hat{g}_n$) of the noise component from the acoustic signal (y(n)) for the time interval;
- Deriving an estimate of the fixed gain ($\hat{g}_s(m)$) of the speech component by subtracting said fixed gain of the noise component from the fixed gain ($g_y(m)$) of the acoustic signal for the time interval.

FIG 4




EP 1 521 243 A1

## Description

### Field of the Invention

**[0001]** The invention refers to a speech coding method applying noise reduction

**[0002]** For over forty years noise reduction methods have been developed in speech processing. Most of the methods are performed in the frequency domain. They commonly comprise three major components:

a) a spectral analysis/synthesis system (typically a short-term windowed FFT (Fast Fourier Transform),
b) IFFT (Inverse Fast Fourier Transform), a noise estimation procedure, and c) a spectral gain computation according to a suppression rule, which is used for suppressing the noise.

**[0003]** The suppression rule modifies only the spectral amplitude, not the phase. It has been shown, that there is no need to modify the phase in speech enhancement processing. Nevertheless, this approximation is only valid for a Signal to Noise Ratio (SNR) greater than 6dB. However, this condition is supposed to be satisfied in the majority of the noise reduction algorithms.

**[0004]** Methods for spectral weighting noise reduction are often based on the following hypothesis:

- The noise is additive (i.e. $y(t)=s(t)+n(t)$), uncorrelated with the speech signal and locally stationary. s and y represent the clean and the noisy speech signal respectively.
- There are silence periods in the speech signal.
- The human auditory system is not sensible to the received speech phase.

**[0005]** A scheme of a treatment of a speech signal with noise reduction is depicted in Fig. 1. The speech component s(p), where p denotes a time interval is superimposed with a noise component n(p). This results in the total signal y (p). The total signal y(p) undergoes a FFT. The result are Fourier components $Y(p, f_k)$, where $f_k$ denotes a quantized frequency. Now the noise reduction NR is applied, thus producing modified Fouriercomponents S(p, $\hat{S}(p,f_k)$. This leads after an IFFT to a clean speech signal estimate $\hat{s}(p)$.

**[0006]** A problem of any spectral weighting noise reduction method is its computational complexity, e.g. if the following steps have to be performed successively:

a) decoding
b) FFT analysis
c) Speech enhancement, e.g. noise reduction
d) Inverse FFT analysis
e) encoding

**[0007]** Thereby the above list is typical for classical noise reduction occurring in a communications network

**[0008]** Based on the foregoing description it is an object of the invention to provide a possibility of a noise reduction method in speech processing systems that can be implemented with a low computational effort.

**[0009]** This object is solved by the subject matter disclosed in the independent claims. Advantageous embodiments of the present invention will be presented in the dependent claims.

**[0010]** In a method for transmitting speech data said speech data are encoded by using an analysis through synthesis method. For the analysis through synthesis a synthesised signal is produced for approximating the original signal. The production of the synthesised signal is performed by using at least a fixed codebook with a respective fixed gain and optionally an adaptive codebook and a adaptive gain. The entries of the codebook and the gain are chosen such, that the synthesised signal resembles the original signal.

**[0011]** Parameters describing these quantities will be transmitted from a sender to a receiver, e.g. from a near-end speaker to a far-end speaker or vice versa.

**[0012]** The invention is based on the idea of modifying the fixed gain determined for the signal containing a noise component and a speech component. Objective of this modification is to obtain a useful estimate of the fixed gain of the speech component or clean signal.

**[0013]** The modification is done by subtraction of an estimate of the fixed gain of the noise component. The fixed gain of the noise component may be derived from an analysis of the power of the signal in a predetermined time window.

**[0014]** One advantage of this procedure is its low computational complexity, particularly if the speech enhancement through noise reduction is done independently from an encoding / decoding unit, e.g. in a certain position within a network, where according to a noise reduction method in the time domain all the steps of decoding, FFT, speech enhancement , IFFT and encoding would have to be performed one after the other. This is not necessary for a noise reduction method according based on modification of parameters

**[0015]** Another advantage is that by using the parameters for any modification, a repeated encoding and decoding process, the so called "tandeming" can be avoided, because the modification takes place in the parameter itself. Any tandeming decreases the speech quality. Furthermore the delay due to the additional encoding/decoding, which is e. g. in GSM typically 5 ms can be avoided.

**[0016]** Thus the parameters, which are actually transmitted do not need to be transformed in a signal for applying the speech reduction. The procedure is furthermore also applicable within a communications network.

**[0017]** An encoding apparatus set up for performing the above described encoding method includes at least a processing unit. The encoding apparatus may be part of a communications device, e.g. a cellular phone or it may be also situated in a communication network or a component thereof.

**[0018]** In the following the invention will be described by means of preferred embodiments with reference to the accompanying drawings in which:

Fig. 1: Scheme of a noise reduction in the frequency domain

Fig. 2: shows schematically the function of the AMR encoder;

Fig. 3: shows schematically the function of the AMR decoder;

Fig. 4: Scheme of a noise reduction method in the parameter domain

**1. Function of a encoder (Fig.2)**

**[0019]** First the function of a speech codec is described by an special implementation of an CELP based codec, the AMR (Adaptive Multirate Codec) codec. The codec consists of a multi-rate, that is, the AMR codec can switch between the following bit rates: 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15 and 4.75 kbit/s, speech codec, a source-controlled rate scheme including a Voice Activity Detection (VAD), a comfort noise generation system and an error concealment mechanism to compensate the effects of transmission errors.

**[0020]** Fig. 2 shows the scheme of the AMR encoder. It uses a LTP (long term prediction) filter. It is transformed to an equivalent structure called adaptive codebook. This codebook saves former LPC filtered excitation signals. Instead of subtracting a long-term prediction as the LTP filter does, an adaptive codebook search is done to get an excitation vector from further LPC filtered speech samples. The amplitude of this excitation is adjusted by a gain factor $g_a$.

**[0021]** The encoding of the speech is described now with reference to the numbers given in Fig. 2

1. The speech signal is processed block-wise and thus partitioned into frames and sub-frames. Each frame is 20 ms long (160 samples at 8 kHz sampling frequency) and is divided into 4 sub-frames of equal length.
2. LPC analysis of a Hamming-windowed frame.
3. Because of stability reasons, the LPC filter coefficients are transformed to Line Spectrum Frequencies (LSF). Afterwards these coefficients are quantized in order to save bit rate. This step and the previous are done once per frame (except in 12.2 kbit/s mode; the LPC coefficients are calculated and quantised twice per frame) whereas the steps 4 - 9 are performed on sub-frame basis.
4. The sub-frames are filtered by a LPC filter with re-transformed and quantised LSF coefficients. Additionally the filter is modified to improve the subjective listening quality.
5. As the encoding is processed block by block, the decaying part of the filter, which is longer than the block length, has to be considered by processing the next sub-frame. In order to speed up the minimization of the residual power described in the following, the zero impulse response of the synthesis filter excited by previous sub-frames is subtracted.
6. The power of the LPC filtered error signal e(n) depends on four variables: the excitation of the adaptive codebook, the excitation of the fixed codebook and the respective gain factors $g_a$ and $g_f$. In order to find the global minimum of the power of the residual signal and as no closed solution of this problem exists, all possible combinations of these four parameters have to be tested experimentally. As the minimization is hence too complex, the problem is divided into subproblems. This results in a suboptimal solution, of course. First the adaptive codebook is searched to get the optimal lag L and gain factor $g_{a,L}$. Afterwards the optimal excitation scaled with the optimal gain factor is synthesis-filtered and subtracted from the target signal. This adaptive codebook search accords to a LTP filtering.
7. In a second step of the minimization problem the fixed codebook is searched. The search is equivalent to the previous adaptive codebook search. I.e. it is looked for the codebook vector that minimizes the error criteria. Afterwards the optimal fixed gain is determined. The resulting coding parameters are the index of the fixed codebook vector J and the optimal gain factor $g_{f,J}$.

8. The scaling factors of the codebooks are quantized jointly (except in 12.2 kbit/s mode - both gains are quantized scalar), resulting in a quantization index, which is also transmitted to the decoder.
9. Completing the processing of the sub-frame, the optimal excitation signal is computed and saved in the adaptive codebook. The synthesis filter states are also saved so that this decaying part can be subtracted in the next sub-frame.

**2. Function of a decoder (Fig. 3)**

**[0022]** Now the decoder is described in reference with Fig. 3. As shown in the previous section, the encoder transforms the speech signal to parameters which describe the speech. We will refer to these parameters, namely the LSF (or LPC) coefficients, the lag of the adaptive codebook, the index of the fixed codebook and the codebook gains, as "speech coding parameters". The domain will be called "(speech) codec parameter domain" and the signals of this domain are subscripted with frame index $k$.

**[0023]** Fig. 3 shows the signal flow of the decoder. The decoder receives the speech coding parameters and computes the excitation signal of the synthesis filter. This excitation signal is the sum of the excitations of the fixed and adaptive codebook scaled with their respective gain factors.

**[0024]** After the synthesis-filtering is performed, the speech signal is post-processed.

**3. Embodiment of a noise reduction rule**

**[0025]** Now an embodiment is described, where the fixed codebook gain of a CELP codec through a certain noise reduction rule is modified such, that the processed fixed codebook is assumed to be noise free. Therfor the following steps are performed, which lead to a less noisy signal after processing.

a) a noisy signal y(t) is coded through a CELP (Code excited linear prediction) codec, e.g. the AMR codec.
b) There the signal is described or coded through the so named 'parameters', i.e. the fixed code book entry, the fixed code book gain, the adaptive code book entry, the adaptive codebook gain, the LPC coefficients etc..
c) With a special processing the fixed code book gain $g_y(m)$ of the signal is extracted from these parameters.
d) A noise reduction is applied to $g_y(m)$.

d1) Accordingly an estimation $\hat{g}_n(m)$ of the noise fixed gain is needed.
d2) Furthermore a reduction rule is required to be applied on the noisy fixed gain $g_y(m)$. An 'unnoisy' or clean fixed gain, i.e an estimation of the speech gain $\hat{g}_s(m)$, is thus obtained.

e) The coded signal is then recomputed by interchanging the noisy fixed gain $g_y(m)$ by the estimation of the speech gain $\hat{g}_s(m)$ and letting the other codec parameters unchanged. The resulting set of codec parameters are assumed to code a clean signal.
f) Optionally a postfilter is applied in order to control the noise reduction rule and to avoid artefacts stemming from the reduction rule.
g) If the encoded, in the above described way modified signal is decoded, a clean signal in the time domain is achieved.

**[0026]** This procedure is depicted schematically in Fig. 4. In an encoder a (total) signal containing clean speech or a speech component and a noise component is encoded. By the encoding process a fixed gain $g_y(m)$ of the total signal is calculated. This fixed gain $g_y(m)$ of the total signal is subject to a gain modification which bases on a noise gain estimation. By the noise gain estimation an estimate of the fixed gain $\hat{g}_n(m)$ is determined, which is used for the gain modification. The result of the gain modification is an estimate of the fixed gain $\hat{g}_s(m)$ of the clean speech or the speech component. This parameter is transmitted from a sender to a receiver. At the receiver side it is decoded. This procedure, especially the noise reduction rule, will now be described in detail:

a) Gain subtraction

**[0027]** One possibility to achieve parameters representing the fixed signal is gain subtraction. Assuming that the fixed codebook gain $g_y(m)$ from the noisy speech is the sum of the clean fixed codebook gain from the clean speech and the noisy fixed codebook gain from the noise, the fixed codebook gain is modified accordingly:

$$\hat{g}_s(m)=g_y(m)-\hat{g}_n(m),$$

where m denotes a time interval, e.g. a frame or a subframe, $\hat{g}_n(m)$ the estimate of the noise component and $\hat{g}_s(m)$ the estimate of the clean codebook gain. It will be described in the next section in reference with a different embodiment, how the estimate of fixed gain $\hat{g}_n(m)$ of the noise component can be calculated.

b) estimation of $\hat{g}_n$

**[0028]** For the above described subtraction the knowledge of the estimate $\hat{g}_n$ of the noise fixed gain is required. The estimation of $\hat{g}_n$ is based on the principle of minimum statistics, wherein the short-term minimum of the estimation of the noisy signal power P(m) is searched:

$$P(m) = \alpha(m).P(m\text{-}1)+(1\text{-}\alpha(m)).g_y^2(m),$$

with

$$\alpha(m) = \max\left[0.3, \frac{\alpha_{\max}}{1+\left(\dfrac{P(m-1)}{\hat{\sigma}_N^2(m-1)}-1\right)^2}\right],$$

where $\hat{\sigma}_N^2$ is the estimate of the noise power found by using the minimal value the noisy signal power *P* on a window of length *D*. As a noisy speech signal contains speech and noise, the minimum value, which is present over a certain time period, and which occurs e.g. in speech pauses represents the noise power $\hat{\sigma}_N^2$.
$\alpha_{\max}$ is a constant, e.g. an advantageous embodiment uses $\alpha_{\max}$ =0.96.

**[0029]** To reduce the number of required comparisons for estimating the noise power $\hat{\sigma}_N^2$, that window of length D is divided in *U* sub-windows of length *V*. The minimum value in the window of length D is the minimum of the set of minimums on each subwindow. A buffer, Min_I of U elements contains the set of minimums from the last U sub-windows. It is renewed each time that V values of P are computed. The oldest element of the buffer is deleted and replaced by the minimum of the last V values of P. The minimum on the window of length D, $\hat{\sigma}_N^2$ for each sub-frame m is the minimum between the minimum of the buffer and the last value of P computed. $\hat{\sigma}_N^2$ can be increased by a gain parameter *omin* to compensate the bias of the estimation. A bias might be due to a continued overestimating of the noise, e.g. if a continually present murmuring is considered as noise only.

**[0030]** The value of $\hat{g}_n$ is finally given by:

$$\hat{g}_n = \sqrt{\hat{\sigma}_N^2(m)}$$

d) postfiltering to control the noise reduction

**[0031]** The noise reduction, as it has been described above, may cause some artefacts during the voice activity periods, e.g. that the speech signal is attenuated due to an overestimation of the noise component

## Claims

1. Method for encoding an acoustic signal (y(n)) containing a speech component and a noise component by using an analysis through synthesis method, wherein for encoding the acoustic signal a synthesised signal is compared with the acoustic signal for a time interval, said synthesised signal being described by using a fixed codebook and an associated fixed gain, comprising the steps:

   a) Determining a fixed gain ($g_y(m)$) of the acoustic signal (y(n)) for the time interval;
   b) Extracting an estimated fixed gain ($\hat{g}_n$) of the noise component from the acoustic signal (y(n)) for the time interval;

c) Deriving an estimate of the fixed gain ($\hat{g}_s(m)$) of the speech component by subtracting said fixed gain of the noise component from the fixed gain ($g_y(m)$) of the acoustic signal for the time interval.

**2.** Method according to claim 1, wherein the synthesised signal is further described by an adaptive codebook and an associated adaptive gain.

**3.** Method according to any of the previous claims, wherein the extracting in step b) is done by searching the minimum value of a power of the signal (y(n)) in a time interval, a part of a time interval or a set of time intervals.

**4.** Method according to any of the previous claims, wherein said time intervalls are frames or subframes.

**5.** Noise reducing apparatus with a processing unit set up for performing a method according to any of the claims 1 to 4.

**6.** Communications device, in particular a mobile phone with a noise reducing apparatus according to claim 5.

**7.** Communications network with a noise reducing apparatus according to claim 5.

FIG 1

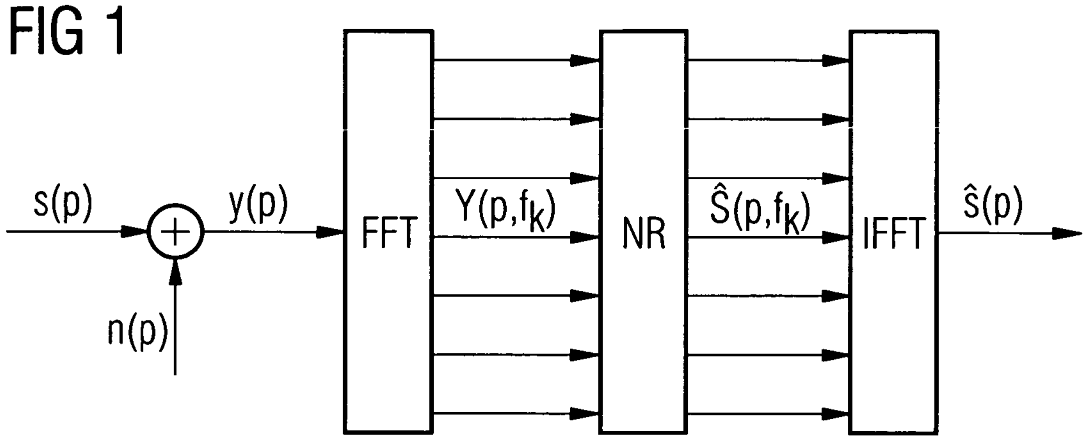


FIG 2

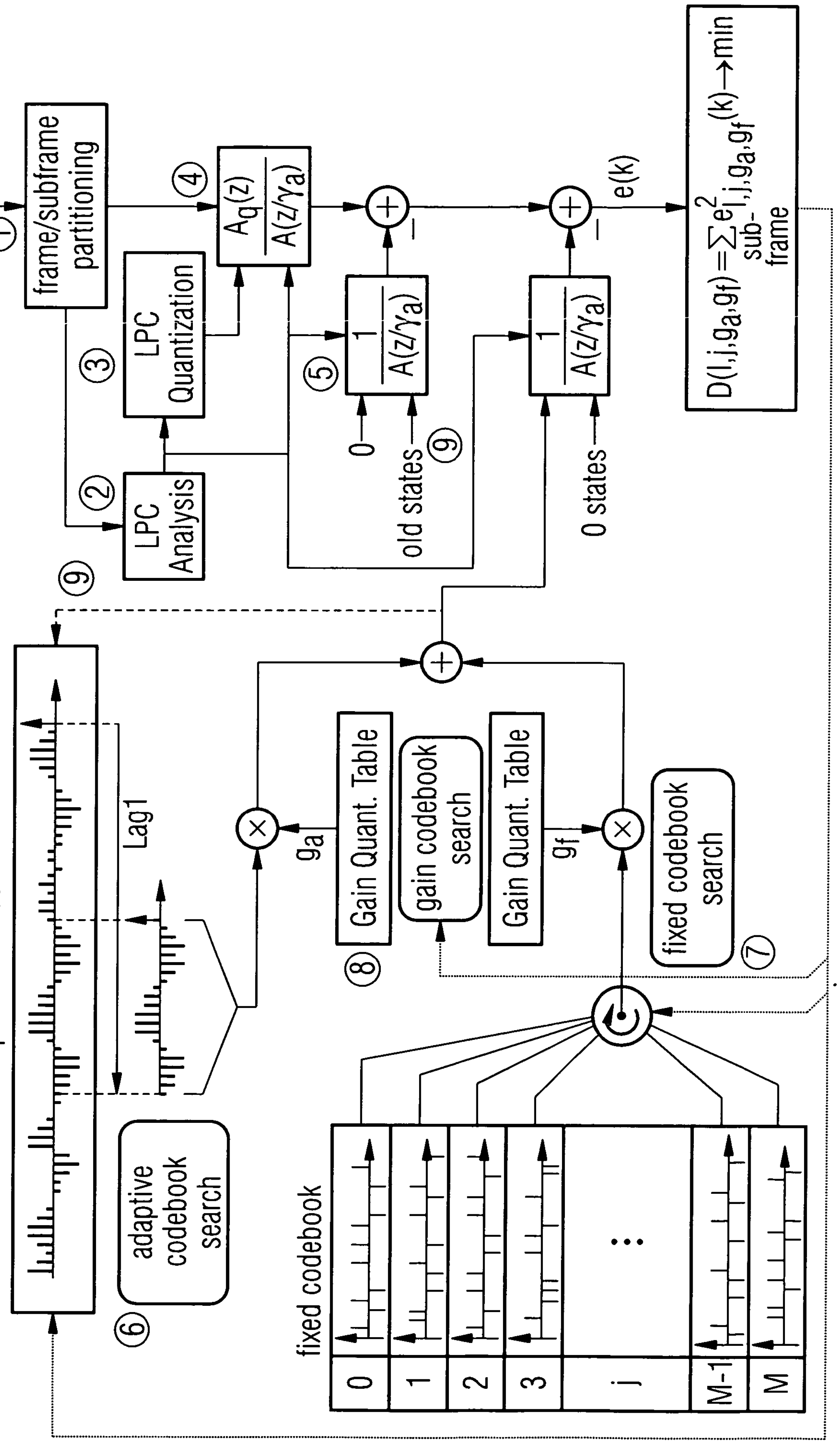

Speech input
1
frame/subframe partitioning
2
LPC Analysis
3
LPC Quantization
4
$\frac{A_q(z)}{A(z/\gamma_a)}$
5
$\frac{1}{A(z/\gamma_a)}$
0
old states
9
0 states
$\frac{1}{A(z/\gamma_a)}$
e(k)
$D(l,j,g_a,g_f)=\sum_{\text{sub-frame}} e^2_{l,j,g_a,g_f}(k)\rightarrow \min$
9
adaptive codebook
Lag1
adaptive codebook search
6
$g_a$
Gain Quant. Table
gain codebook search
8
Gain Quant. Table
$g_f$
fixed codebook search
7
fixed codebook
0
1
2
3
j
M-1
M

FIG 3

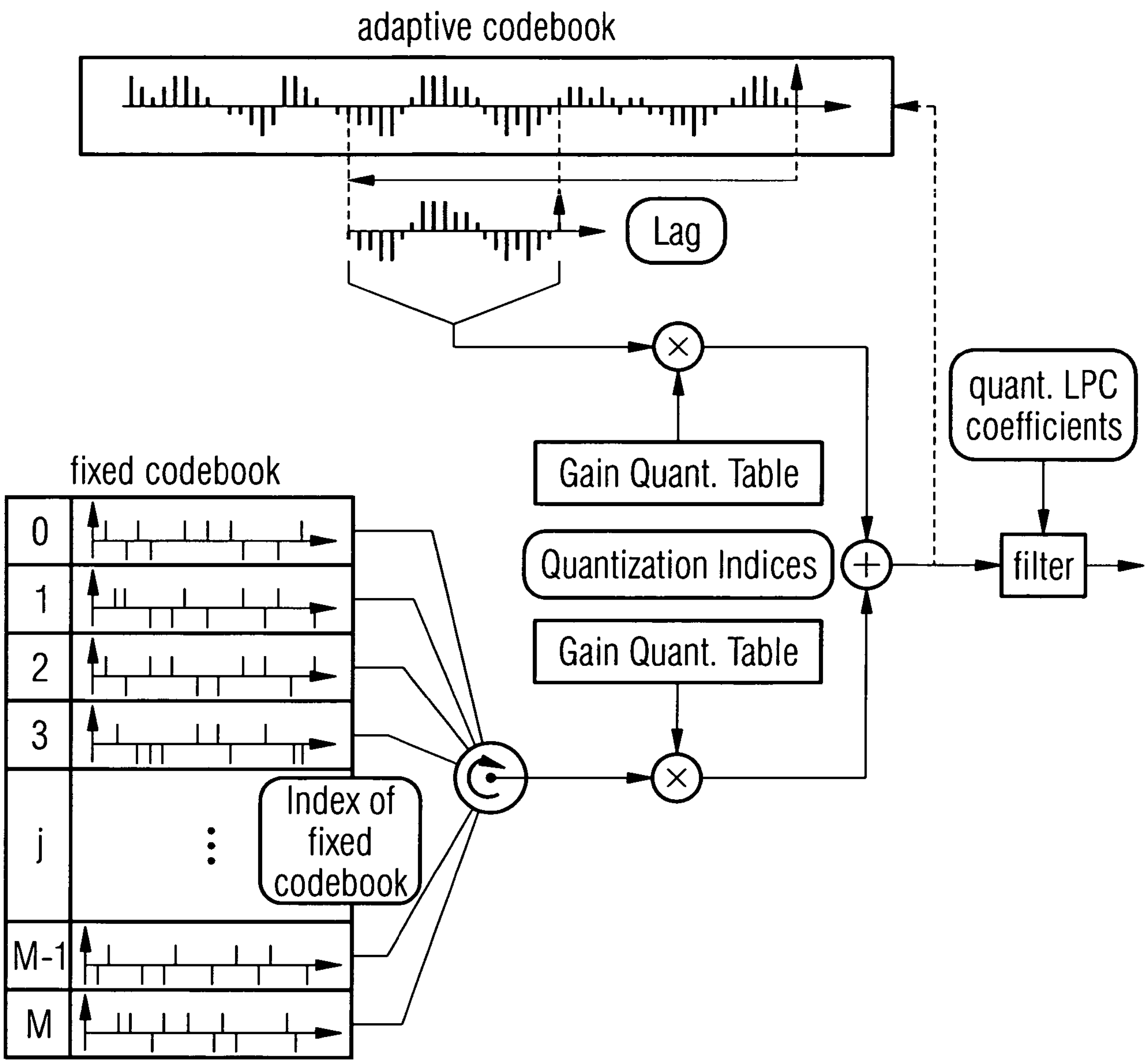
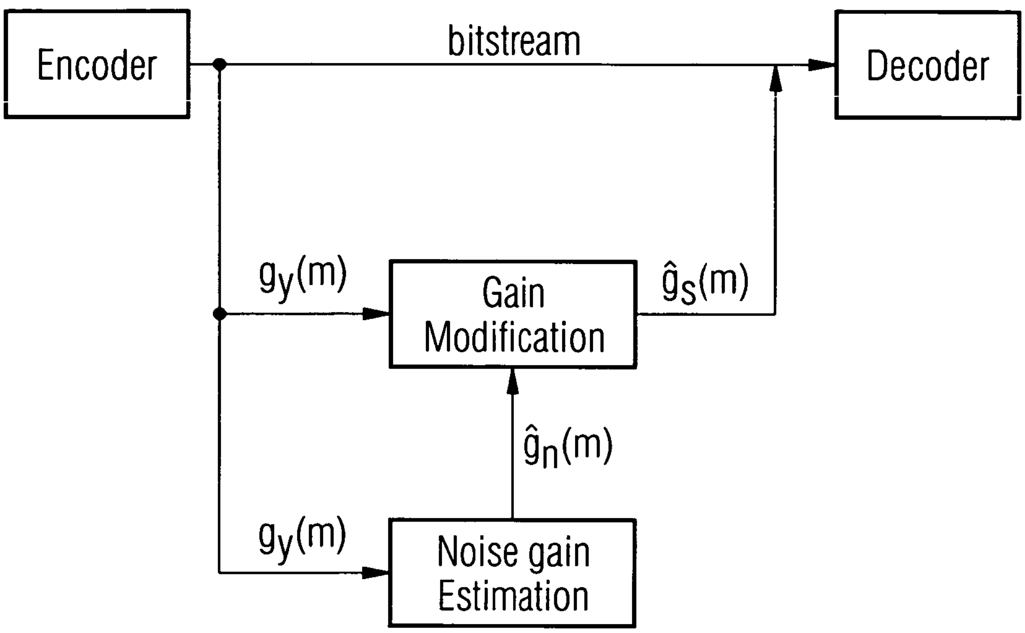

adaptive codebook
Lag
quant. LPC
coefficients
fixed codebook
Gain Quant. Table
Quantization Indices
filter
Gain Quant. Table
0
1
2
3
j
Index of
fixed
codebook
M-1
M

FIG 4

Scheme
noise
reduction
on codec



European Patent Office

# EUROPEAN SEARCH REPORT

Application Number
EP 03 02 2251

## DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document with indication, where appropriate, of relevant passages | Relevant to claim | CLASSIFICATION OF THE APPLICATION (Int.Cl.7) |
|---|---|---|---|
| X<br><br><br><br><br><br><br><br><br><br>Y | CHANDRAN R ET AL: "COMPRESSED DOMAIN NOISE REDUCTION AND ECHO SUPPRESSION FOR NETWORK SPEECH ENHANCEMENT"<br>PROCEEDINGS OF THE 43RD. IEEE MIDWEST SYMPOSIUM ON CIRCUITS AND SYSTEMS. MWSCAS 2000. LANSING, MI. NEW YORK, NY: IEEE, US, vol. 1 OF 3, 8 - 11 August 2000, pages 10-13, XP002951730<br>ISBN: 0-7803-6476-7<br>* the whole document *<br>--- | 1,3-7<br><br><br><br><br><br><br><br><br><br>2 | G10L21/02<br>G10L19/04 |
| Y | US 2002/184010 A1 (ERIKSSON ET AL)<br>5 December 2002 (2002-12-05)<br>* abstract *<br>* page 1, paragraph 4 - paragraph 6 *<br>* page 2, paragraph 21 - paragraph 33 *<br>* page 3, paragraph 46 - paragraph 57 *<br>* claims 1-5; figures 3,8 *<br>--- | 2 | |
| A | EP 1 301 018 A (CIT ALCATEL)<br>9 April 2003 (2003-04-09)<br>* page 3, paragraph 13 - paragraph 22 *<br>* page 5, paragraph 45 - paragraph 52 *<br>* figures 2-4 *<br>--- | 1-7 | TECHNICAL FIELDS SEARCHED (Int.Cl.7)<br>G10L |
| A | LIM J S ET AL: "ENHANCEMENT AND BANDWIDTH COMPRESSION OF NOISY SPEECH"<br>PROCEEDINGS OF THE IEEE, IEEE. NEW YORK, US,<br>vol. 67, no. 12, December 1979 (1979-12), pages 1586-1604, XP000891496<br>ISSN: 0018-9219<br>* abstract *<br>----- | 1 | |

The present search report has been drawn up for all claims

| Place of search | Date of completion of the search | Examiner |
|---|---|---|
| THE HAGUE | 8 March 2004 | Santos Luque, R |

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

1

EPO FORM 1503 03.82 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO.** EP 03 02 2251

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

08-03-2004

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| US 2002184010 A1 | 05-12-2002 | SE 521693 C2 | 25-11-2003 |
| | | GB 2390790 A | 14-01-2004 |
| | | SE 0102519 A | 01-10-2002 |
| | | WO 02080149 A1 | 10-10-2002 |
| EP 1301018 A | 09-04-2003 | EP 1301018 A1 | 09-04-2003 |
| | | US 2003065507 A1 | 03-04-2003 |

EPO FORM P0459

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