Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2887009 C 2019/12/17

(11)(21) 2 887 009

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2013/07/31
(87) Date publication PCT/PCT Publication Date: 2014/04/10
(45) Date de délivrance/Issue Date: 2019/12/17
(85) Entrée phase nationale/National Entry: 2015/04/01
(86) N° demande PCT/PCT Application No.: EP 2013/066074
(87) N° publication PCT/PCT Publication No.: 2014/053261
(30) Priorité/Priority: 2012/10/05 (US61/710,137)

(51) Cl.Int./Int.Cl. *G10L 19/107* (2013.01)

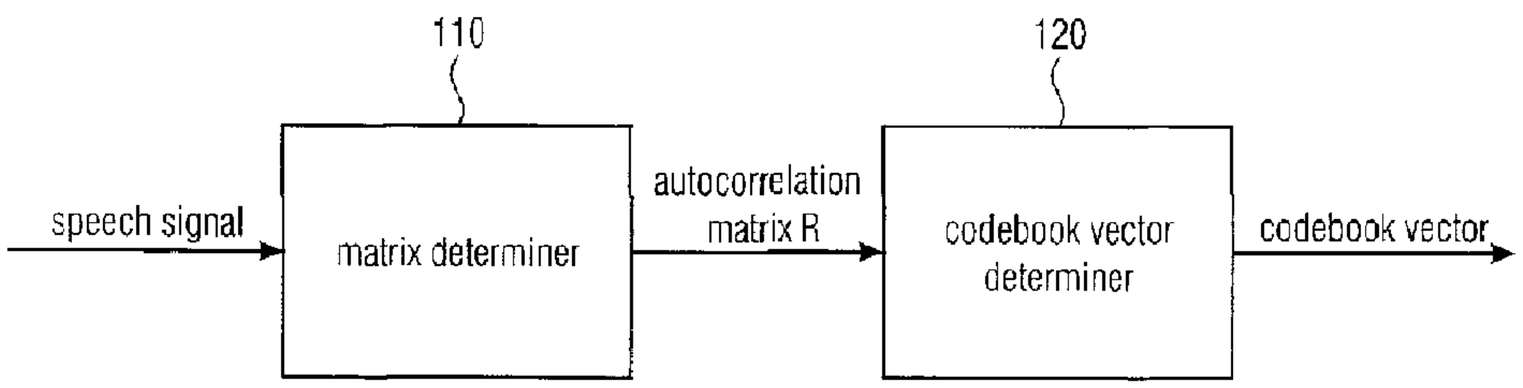
(72) Inventeurs/Inventors:
BACKSTROM, TOM, DE;
MULTRUS, MARKUS, DE;
FUCHS, GUILLAUME, DE;
HELMRICH, CHRISTIAN, DE;
DIETZ, MARTIN, DE

(73) Propriétaire/Owner:
FRAUNHOFER-GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG E.V., DE

(74) Agent: BORDEN LADNER GERVAIS LLP

(54) Titre : APPAREIL POUR CODER UN SIGNAL DE PAROLE EMPLOYANT ACELP DANS LE DOMAINE D'AUTOCORRELATION

(54) Title: AN APPARATUS FOR ENCODING A SPEECH SIGNAL EMPLOYING ACELP IN THE AUTOCORRELATION DOMAIN



(57) Abrégé/Abstract:
An apparatus for encoding a speech signal by determining a codebook vector of a speech coding algorithm is provided. The apparatus comprises a matrix determiner (110) for determining an autocorrelation matrix R, and a codebook vector determiner (120) for determining the codebook vector depending on the autocorrelation matrix R. The matrix determiner (110) is configured to determine the autocorrelation matrix R by determining vector coefficients of a vector r, wherein the autocorrelation matrix R comprises a plurality of rows and a plurality of columns, wherein the vector r indicates one of the columns or one of the rows of the autocorrelation matrix R, wherein R(i, j) = r(|i-j|), wherein R(i, j) indicates the coefficients of the autocorrelation matrix R, wherein i is a first index indicating one of a plurality of rows of the autocorrelation matrix R, and wherein j is a second index indicating one of the plurality of columns of the autocorrelation matrix R.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 April 2014 (10.04.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/053261 A1

(51) International Patent Classification:
G10L 19/107 (2013.01)

(21) International Application Number:
PCT/EP2013/066074

(22) International Filing Date:
31 July 2013 (31.07.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/710,137 5 October 2012 (05.10.2012) US

(71) Applicant: **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Hansastraße 27c, 80686 München (DE).

(72) Inventors: **BÄCKSTRÖM, Tom**; Bauerngasse 8-12, 90443 Nürnberg (DE). **MULTRUS, Markus**; Etzlaubweg 7, 90469 Nürnberg (DE). **FUCHS, Guillaume**; Fürther Straße 17, 91058 Erlangen (DE). **HELMRICH, Christian**; Hauptstraße 68, 91054 Erlangen (DE). **DIETZ, Martin**; Deutschherrnstraße 37, 90429 Nürnberg (DE).

(74) Agents: **ZINKLER, Franz** et al.; 246, 82043 Pullach (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

— *with international search report (Art. 21(3))*

(54) Title: AN APPARATUS FOR ENCODING A SPEECH SIGNAL EMPLOYING ACELP IN THE AUTOCORRELATION DOMAIN

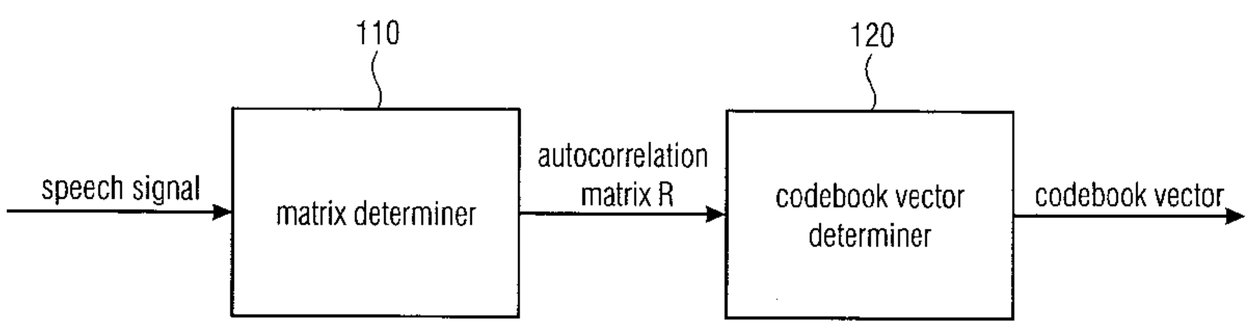


FIG 1

(57) Abstract: An apparatus for encoding a speech signal by determining a codebook vector of a speech coding algorithm is provided. The apparatus comprises a matrix determiner (110) for determining an autocorrelation matrix *R*, and a codebook vector determiner (120) for determining the codebook vector depending on the autocorrelation matrix *R*. The matrix determiner (110) is configured to determine the autocorrelation matrix *R* by determining vector coefficients of a vector *r*, wherein the autocorrelation matrix *R* comprises a plurality of rows and a plurality of columns, wherein the vector *r* indicates one of the columns or one of the rows of the autocorrelation matrix *R*, wherein *R*(i, j) = *r*(Ii—jI), wherein *R*(*i*, *j*) indicates the coefficients of the autocorrelation matrix *R*, wherein *i* is a first index indicating one of a plurality of rows of the autocorrelation matrix *R*, and wherein *j* is a second index indicating one of the plurality of columns of the autocorrelation matrix *R*.

WO 2014/053261 A1

An Apparatus for Encoding a Speech Signal employing ACELP in the Autocorrelation Domain

Description

The present invention relates to audio signal coding, and, in particular, to an apparatus for encoding a speech signal employing ACELP in the autocorrelation domain.

In speech coding by Code-Excited Linear Prediction (CELP), the spectral envelope (or equivalently, short-time time-structure) of the speech signal is described by a linear predictive (LP) model and the prediction residual is modelled by a long-time predictor (LTP, also known as the adaptive codebook) and a residual signal represented by a codebook (also known as the fixed codebook). The latter, the fixed codebook, is generally applied as an algebraic codebook, where the codebook is represented by an algebraic formula or algorithm, whereby there is no need to store the whole codebook, but only the algorithm, while simultaneously allowing for a fast search algorithm. CELP codecs applying an algebraic codebook for the residual are known as Algebraic Code-Excited Linear Prediction (ACELP) codecs (see [1], [2], [3], 4]) .

In speech coding, employing an algebraic residual codebook is the approach of choice in main stream codecs such as [17], [13], [18]. ACELP is based on modeling the spectral envelope by a linear predictive (LP) filter, the fundamental frequency of voiced sounds by a long time predictor (LTP) and the prediction residual by an algebraic codebook. The LTP and algebraic codebook parameters are optimized by a least squares algorithm in a perceptual domain, where the perceptual domain is specified by a filter.

The computationally most complex part of ACELP-type algorithms, the bottleneck, is optimization of the residual codebook. The only currently known optimal algorithm would be an exhaustive search of a size N^p space for every sub-frame, where at every point, an evaluation of $\mathcal{O}(N^2)$ complexity is required. Since typical values are sub-frame length N = 64 (i.e. 5ms) with p = 8 pulses, this implies more than 10^{20} operations per second. Clearly this is not a viable option. To stay within the complexity limits set by hardware requirements, codebook optimization approaches have to operate with non-optimal iterative algorithms. Many such algorithms and improvements to the optimization process have been presented in the past, for example [17], [19], [20], [21], [22].

Explicitly, the ACELP optimisation is based on describing the speech signal x(n) as the output of a linear predictive model such that the estimated speech signal is

$$\hat{x}(n) = -\sum_{k=1}^{m} a(k)\hat{x}(n-k) + \hat{e}(k) \quad (1)$$

where a(k) are the LP coefficients and $\hat{e}(k)$ is the residual signal. In vector form, this equation can be expressed as

$$\hat{x} = H\hat{e} \quad (2)$$

where matrix H is defined as the lower triangular Toeplitz convolution matrix with diagonal h(0) and lower diagonals h(1), ..., h(39) and the vector h(k) is the impulse response of the LP model. It should be noted that in this notation the perceptual model (which usually corresponds to a weighted LP model) is omitted, but it is assumed that the perceptual model is included in the impulse response h(k). This omission has no impact on the generality of results, but simplifies notation. The inclusion of the perceptual model is applied as in [1].

The fitness of the model is measured by the squared error. That is,

$$\epsilon^2 = \sum_{k=1}^{N} (x(k) - \hat{x}(k))^2 = (e - \hat{e})^H H^H H (e - \hat{e}). \quad (3)$$

This squared error is used to find the optimal model parameters. Here, it is assumed that the LTP and the pulse codebook are both used to model the vector e. The practical application can be found in the relevant publications (see [1-4]).

In practice, the above measure of fitness can be simplified as follows. Let the matrix $B = H^T H$ comprise the correlations of h(n), let c_k be the k'th fixed codebook vector and set $\hat{e} = g\,c_k$, where g is a gain factor. By assuming that g is chosen optimally, then the codebook is searched by maximizing the search criterion

$$\frac{C_k^2}{E_k} = \frac{(x^T H c_k)^2}{c_k^T B c_k} = \frac{(d^T c_k)^2}{c_k^T B c_k} \quad (4)$$

where $d = H^T x$ is a vector comprising the correlation between the target vector and the impulse response h(n) and superscript T denotes transpose. The vector d and the matrix B

are computed before the codebook search. This formula is commonly used in optimization of both the LTP and the pulse codebook.

Plenty of research has been invested in optimising the usage of the above formula. For example,

1) Only those elements of matrix B are calculated that are actually accessed by the search algorithm. Or:

2) The trial-and-error algorithm of the pulse search is reduced to trying only such codebook vectors which have a high probability of success, based on prior screening (see for example [1,5]).

A practical detail of the ACELP algorithm is related to the concept of zero impulse response (ZIR). The concept appears when considering the original domain synthesis signal in comparison to the synthesised residual. The residual is encoded in blocks corresponding to the frame or sub-frame size. However, when synthesising the original domain signal with the LP model of Equation 1, the fixed length residual will have an infinite length "tail", corresponding to the impulse response of the LP filter. That is, although the residual codebook vector is of finite length, it will have an effect on the synthesis signal far beyond the current frame or sub-frame. The effect of a frame into the future can be calculated by extending the codebook vector with zeros and calculating the synthesis output of Equation 1 for this extended signal. This extension of the synthesised signal is known as the zero impulse response. Then, to take into account the effect of prior frames in encoding the current frame, the ZIR of the prior frame is subtracted from the target of the current frame. In encoding the current frame, thus, only that part of the signal is considered, which was not already modelled by the previous frame.

In practice, the ZIR is taken into account as follows: When a (sub)frame N-1 has been encoded, the quantized residual is extended with zeros to the length of the next (sub)frame N. The extended quantized residual is filtered by the LP to obtain the ZIR of the quantized signal. The ZIR of the quantized signal is then subtracted from the original (not quantized) signal and this modified signal forms the target signal when encoding (sub)frame N. This way, all quantization errors made in (sub)frame N-1 will be taken into account when quantizing (sub)frame N. This practice improves the perceptual quality of the output signal considerably.

However, it would be highly appreciated if further improved concepts for audio coding would be provided.

The object of the present invention is to provide such improved concepts for audio object coding.

An apparatus for encoding a speech signal by determining a codebook vector of a speech coding algorithm is provided. The apparatus comprises a matrix determiner for determining an autocorrelation matrix R, and a codebook vector determiner for determining the codebook vector depending on the autocorrelation matrix R. The matrix determiner is configured to determine the autocorrelation matrix R by determining vector coefficients of a vector r, wherein the autocorrelation matrix R comprises a plurality of rows and a plurality of columns, wherein the vector r indicates one of the columns or one of the rows of the autocorrelation matrix R, wherein $R(i\ ,\ j) = r(|i- j|)$, wherein $R(i,\ j)$ indicates the coefficients of the autocorrelation matrix R, wherein i is a first index indicating one of a plurality of rows of the autocorrelation matrix R, and wherein j is a second index indicating one of the plurality of columns of the autocorrelation matrix R.

The apparatus is configured to use the codebook vector to encode the speech signal. For example, the apparatus may generate the encoded speech signal such that the encoded speech signal comprises a plurality of Linear Prediction coefficients, an indication of the fundamental frequency of voiced sounds (e.g., pitch parameters), and an indication of the codebook vector, e.g, an index of the codebook vector.

Moreover, a decoder for decoding an encoded speech signal being encoded by an apparatus according to the above-described embodiment to obtain a decoded speech signal is provided.

Furthermore a system is provided. The system comprises an apparatus according to the above-described embodiment for encoding an input speech signal to obtain an encoded speech signal. Moreover, the system comprises a decoder according to the above-described embodiment for decoding the encoded speech signal to obtain a decoded speech signal.

Improved concepts for the objective function of the speech coding algorithm ACELP are provided, which take into account not only the effect of the impulse response of the

previous frame to the current frame, but also the effect of the impulse response of the current frame into the next frame, when optimizing parameters of current frame. Some embodiments realize these improvements by changing the correlation matrix, which is central to conventional ACELP optimisation to an autocorrelation matrix, which has Hermitian Toeplitz structure. By employing this structure, it is possible to make ACELP optimisation more efficient in terms of both computational complexity as well as memory requirements. Concurrently, also the perceptual model applied becomes more consistent and interframe dependencies can be avoided to improve performance under the influence of packet-loss.

Speech coding with the ACELP paradigm is based on a least squares algorithm in a perceptual domain, where the perceptual domain is specified by a filter. According to embodiments, the computational complexity of the conventional definition of the least squares problem can be reduced by taking into account the impact of the zero impulse response into the next frame. The provided modifications introduce a Toeplitz structure to a correlation matrix appearing in the objective function, which simplifies the structure and reduces computations. The proposed concepts reduce computational complexity up to 17% without reducing perceptual quality.

Embodiments are based on the finding that by a slight modification of the objective function, complexity in the optimization of the residual codebook can be further reduced. This reduction in complexity comes without reduction in perceptual quality. As an alternative, since ACELP residual optimization is based on iterative search algorithms, with the presented modification, it is possible to increase the number of iterations without an increase in complexity, and in this way obtain an improved perceptual quality.

Both the conventional as well as the modified objective functions model perception and strive to minimize perceptual distortion. However, the optimal solution to the conventional approach is not necessarily optimal with respect to the modified objective function and vice versa. This alone does not mean that one approach would be better than the other, but analytic arguments do show that the modified objective function is more consistent. Specifically, in contrast to the conventional objective function, the provided concepts treat all samples within a sub-frame equally, with consistent and well-defined perceptual and signal models.

In embodiments, the proposed modifications can be applied such that they only change the optimization of the residual codebook. It does therefore not change the bit-stream structure and can be applied in a back-ward compatible manner to existing ACELP codecs.

Moreover, a method for encoding a speech signal by determining a codebook vector of a speech coding algorithm is provided. The method comprises:

- Determining an autocorrelation matrix R. And:

- Determining the codebook vector depending on the autocorrelation matrix R.

Determining an autocorrelation matrix R comprises determining vector coefficients of a vector r. The autocorrelation matrix R comprises a plurality of rows and a plurality of columns. The vector r indicates one of the columns or one of the rows of the autocorrelation matrix R, wherein

$$R(i\,,\,j) = r(|i-j|).$$

$R(i,\ j)$ indicates the coefficients of the autocorrelation matrix R, wherein i is a first index indicating one of a plurality of rows of the autocorrelation matrix R, and wherein j is a second index indicating one of the plurality of columns of the autocorrelation matrix R.

Furthermore, a method for decoding an encoded speech signal being encoded according to the method for encoding a speech signal according to the above-described embodiment to obtain a decoded speech signal is provided.

Moreover, a method is provided. The method comprises:

- Encoding an input speech signal according to the above-described method for encoding a speech signal to obtain an encoded speech signal. And:

- Decoding the encoded speech signal to obtain a decoded speech signal according to the above-described method for decoding a speech signal.

Furthermore, computer programs for implementing the above-described methods when being executed on a computer or signal processor are provided.

In the following, embodiments of the present invention are described in more detail with reference to the figures, in which:

Fig. 1 illustrates an apparatus for encoding a speech signal by determining a codebook vector of a speech coding algorithm according to an embodiment,

Fig. 2 illustrates a decoder according to an embodiment and a decoder, and

Fig. 3 illustrates a system comprising an apparatus for encoding a speech signal according to an embodiment and a decoder.

Fig. 1 illustrates an apparatus for encoding a speech signal by determining a codebook vector of a speech coding algorithm according to an embodiment.

The apparatus comprises a matrix determiner (110) for determining an autocorrelation matrix R, and a codebook vector determiner (120) for determining the codebook vector depending on the autocorrelation matrix R.

The matrix determiner (110) is configured to determine the autocorrelation matrix R by determining vector coefficients of a vector r.

The autocorrelation matrix R comprises a plurality of rows and a plurality of columns, wherein the vector r indicates one of the columns or one of the rows of the autocorrelation matrix R, wherein $R(i, j) = r(|i-j|)$.

$R(i, j)$ indicates the coefficients of the autocorrelation matrix R, wherein i is a first index indicating one of a plurality of rows of the autocorrelation matrix R, and wherein j is a second index indicating one of the plurality of columns of the autocorrelation matrix R.

The apparatus is configured to use the codebook vector to encode the speech signal. For example, the apparatus may generate the encoded speech signal such that the encoded speech signal comprises a plurality of Linear Prediction coefficients, an indication of the fundamental frequency of voiced sounds (e.g. pitch parameters), and an indication of the codebook vector.

For example, according to a particular embodiment for encoding a speech signal, the apparatus may be configured to determine a plurality of linear predictive coefficients (a(k)) depending on the speech signal. Moreover, the apparatus is configured to determine a residual signal depending on the plurality of linear predictive coefficients (a(k)).

Furthermore, the matrix determiner 110 may be configured to determine the autocorrelation matrix R depending on the residual signal.

In the following, some further embodiments of the present invention are described.

Returning to equations 3 and 4, wherein Equation 3 defines a squared error indicating a fitness of the perceptual model as:

$$\epsilon^2 = \sum_{k=1}^{N} (x(k) - \hat{x}(k))^2 = (e - \hat{e})^H H^H H (e - \hat{e}), \quad (3)$$

and wherein Equation 4

$$\frac{C_k^2}{E_k} = \frac{(x^T H c_k)^2}{c_k^T B c_k} = \frac{(d^T c_k)^2}{c_k^T B c_k} \quad (4).$$

indicates the search criterion, which is to be maximized.

The ACELP algorithm is centred around Equation 4, which in turn is based on Equation 3.

Embodiments are based on the finding that analysis of these equations reveals that the quantized residual values e(k) have a very different effect on the error energy ϵ^2 depending on the index k. For example, when considering the indices k=1 and k=N, if the only non-zero value of the residual codebook would appear at k=1, then the error energy ϵ^2 results to:

$$\epsilon_1^2 = \sum_{k=1}^{N} (x(k) - e(1) h(k))^2 \quad (5)$$

while for k=N, the error energy ϵ^2 results to:

$$\epsilon_N^2 = (x(N) - e(N) h(1))^2 + \sum_{k=1}^{N-1} (x(k))^2. \quad (6)$$

In other words, e(1) is weighted with the impulse response h(k) on the range 1 to N, while e(N) is weighted with only h(1). In terms of spectral weighting, this means that each e(k) is weighted with a different spectral weighting function, such that, in the extreme, e(N) is

linearly-weighted. From a perceptual modelling perspective, it would make sense to apply the same perceptual weight for all samples within a frame. Equation 3 should thus be extended such that it takes into account the ZIR into the next frame. It should be noticed that here, inter alia, the difference to prior art is that both the ZIR from the previous frame and also the ZIR into the next frame are taken into account.

Let e(k) be the original, unquantized residual and $\hat{e}(k)$ the quantised residual. Furthermore, let both residuals be non-zero in the range 1 to N and zero elsewhere. Then

$$\begin{aligned} x(n) &= -\sum_{k=1}^{m} a(k)x(n-k) + e(n) = \sum_{k=1}^{\infty} e(n-k)h(k) \\ \hat{x}(n) &= -\sum_{k=1}^{m} a(k)\hat{x}(n-k) + \hat{e}(n) = \sum_{k=1}^{\infty} \hat{e}(n-k)h(k) \end{aligned} \quad (7)$$

Equivalently, the same relationships in matrix form can be expressed as:

$$\begin{aligned} x &= \tilde{H}\, e \\ \hat{x} &= \tilde{H}\, \hat{e} \end{aligned} \quad (8)$$

where $\tilde{H}$ is the infinite dimensional convolution matrix corresponding to the impulse response h(k). Inserting into Equation 3 yields

$$\epsilon^2 = \|\tilde{H}\, e - \tilde{H}\, \hat{e}\|^2 = (e-\hat{e})^T \tilde{H}^T \tilde{H}\, (e-\hat{e}) = (e-\hat{e})^T R (e-\hat{e}) \quad (9)$$

where $R = \tilde{H}^T \tilde{H}$ is the finite size, Hermitian Toeplitz matrix corresponding to the autocorrelation of h(n). By a similar derivation as for Equation 4, the objective function is obtained:

$$\frac{(e^T R\, \hat{e})^2}{(\hat{e}^T R\, \hat{e})} = \frac{(d^T e)^2}{(\hat{e}^T R\, \hat{e})}. \quad (10)$$

This objective function is very similar to Equation 4. The main difference is that instead of the correlation matrix B, here a Hermitian Toeplitz matrix R is in the denominator.

As explained above, this novel formulation has the benefit that all samples of the residual e within a frame will receive the same perceptual weighting. However, importantly, this formulation introduces considerable benefits to computational complexity and memory requirements as well. Since R is a Hermitian Toeplitz matrix, the first column r(0)..r(N-1)

defines the matrix completely. In other words, instead of storing the complete NxN matrix, it is sufficient to store only the Nx1 vector r(k), thus yielding a considerable saving in memory allocation. Moreover, computational complexity is also reduced since it is not necessary to determine all NxN elements, but only the first Nx1 column. Also indexing within the matrix is simple, since the element (i,j) can be found by $R(i,j) = r(|i-j|)$.

Since the objective function in Equation 10 is so similar to Equation 4, the structure of the general ACELP can be retained. Specifically, any of the following operations can be performed with either objective function, with only minor modifications to the algorithm:

1. Optimisation of the LTP lag (adaptive codebook)

2. Optimisation of the pulse codebook for modelling the residual (fixed codebook)

3. Optimisation of the gains of LTP and pulses, either separately or jointly

4. Optimisation of any other parameters whose performance can be measured by the squared error of Equation 3.

The only part that has to be modified in conventional ACELP applications is the handling of the correlation matrix B, which is replaced by matrix *R*, as well as the target, which must include the ZIR into the following frame.

Some embodiments employ the concepts of the present invention by, wherever in the ACELP algorithm, where the correlation matrix B appears, it is replaced by the autocorrelation matrix *R*. If all instances of the matrix B are omitted, then calculating its value can be avoided.

For example, the autocorrelation matrix *R* is determined by determining the coefficients of the first column r(0), .., r(N-1) of the autocorrelation matrix R.

The matrix *R* is defined in Equation 9 by $R=H^TH$, whereby its elements $R_{ij}=r(i-j)$ can be calculated through

$$r(k) = h(k) * h(-k) = \sum_l h(l)h(l-k) \qquad (9a)$$

That is, the sequence r(k) is the autocorrelation of h(k).

Often, however, r(k) can be obtained by even more effective means. Specifically, in speech coding standards such as AMR and G.718, the sequence h(k) is the impulse response of a linear predictive filter A(z) filtered by a perceptual weighting function W(z), which is taken to include the pre-emphasis. In other words, h(k) indicates a perceptually weighted impulse response of a linear predictive model.

The filter A(z) is usually estimated from the autocorrelation of the speech signal $r_X(k)$, that is, $r_X(k)$ is already known. Since $H(z) = A^{-1}(u)W(z)$, it follows that the autocorrelation sequence r(k) can be determined by calculating the autocorrelation of w(k) by

$$r_w(k) = w(k) * w(-k) = \sum_i w(l)w(l-k) \qquad (9b)$$

whereby the autocorrelation of h(k) is

$$r(k) = r_x(k) * r_w(k) = \sum_i r_w(l)r_x(l-k). \qquad (9c)$$

Depending on the design of the overall system, these equations may, in some embodiments, be modified accordingly.

A codebook vector of a codebook may then, e.g., be determined based on the autocorrelation matrix *R*. In particular, Equation 10 may, according to some embodiments, be used to determine a codebook vector of the codebook.

In this context, Equation 10 defines the objective function in the form $f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$ which is otherwise the same form as in the speech coding standards AMR and G.718 but such that the matrix *R* now has symmetric Toeplitz structure. The objective function is basically a normalized correlation between the target vector d and the codebook vector $\hat{e}$ and the best possible codebook vector is that, which gives the highest value for the normalized correlation $f(\hat{e})$, e.g., which maximizes the normalized correlation $f(\hat{e})$.

Codebook vectors can thus optimized with the same approaches as in the mentioned standards. Specifically, for example, the very simple algorithm for finding the best algebraic codebook (i.e. the fixed codebook) vector $\hat{e}$ for the residual can be applied, as described below. It should, however, be noted, that significant effort has been invested in

the design of efficient search algorithms (c.f. AMR and G.718), and this search algorithm is only an illustrative example of application.

1. Define an initial codebook vector $\hat{e}_0 = [0, 0 \ldots 0]^T$ and set the number of pulses to $p = 0$.

2. Set the initial codebook quality measure to $\hat{f}_0 = 0$.

3. Set temporary codebook quality measure to $\tilde{f}_p = \hat{f}_{p-1}$.

4. For each position k in the codebook vector

   (i) Increase p by one.

   (ii) If position k already contains a negative pulse, continue to step vii.

   (iii) Create a temporary codebook vector $\varepsilon_p^+ = \hat{e}_{p-1}$ and add a positive pulse at position k.

   (iv) Evaluate the quality of the temporary codebook vector by $f(\varepsilon_p^+)$.

   (v) If the temporary codebook vector is better than any of the previous, $f(\varepsilon_p^+) > \tilde{f}_p$, then save this codebook vector, set $\hat{f}_p = f(\varepsilon_p^+)$ and continue to next iteration.

   (vi) If position k already contains a positive pulse, continue to next iteration.

   (vii) Create a temporary codebook vector $\varepsilon_p^- = \hat{e}_{p-1}$ and add a negative pulse at position k.

   (viii) Evaluate the quality of the temporary codebook vector by $f(\varepsilon_p^-)$.

   (ix) If the temporary codebook vector is better than any of the previous, $f(\varepsilon_p^-) > \hat{f}_p$, then save this codebook vector, set $\hat{f}_p = f(\varepsilon_p^-)$ and continue to next iteration.

5. Define the codebook vector $\hat{e}_p$ to be the last (that is, best) of the saved codebook vectors.

6. If the number of pulses p has reached the desired number of pulses, then define the output vector as $\hat{e} = \hat{e}_p$, and stop. Otherwise, continue with step 4.

As already pointed out, compared to conventional ACELP applications, in some embodiments, the target is modified such that it includes the ZIR into the following frame.

Equation 1 describes the linear predictive model used in ACELP-type codecs. The Zero Impulse Response (ZIR, also sometimes known as the Zero Input Response), refers to the output of the linear predictive model when the residual of the current frame (and all future frames) is set to zero. The ZIR can be readily calculated by defining the residual which is zero from position N forward as

$$e_K'(n) = \begin{cases} e(n) & \text{for } n < K \\ 0 & \text{for } n \geq K \end{cases} \qquad (10a)$$

whereby the ZIR can be defined as

$$ZIR_K(n) = \sum_{k=0}^{N} h(k)\, e_K'(n-k). \qquad (10b)$$

By subtracting this ZIR from the input signal, a signal is obtained which depends on the residual only from the current frame forward.

Equivalently, the ZIR can be determined by filtering the past input signal as

$$ZIR_K(n) = \begin{cases} x(n) & \text{for } n < K \\ -\sum_{k=1}^{m} a(k)\, ZIR_K(n-k) & \text{for } n \geq K. \end{cases} \qquad (10c)$$

The input signal where the ZIR has been removed is often known as the target and can be defined for the frame that begins at position K as $d(n) = x(n) - ZIR_K(n)$. This target is in principle exactly equal to the target in the AMR and G.718 standards. When quantizing the signal, the quantized signal $\hat{d}(n)$ is compared to $d(n)$ for the duration of a frame $K \leq n < K + N$.

Conversely, the residual of the current frame has an influence on the following frames, whereby it is useful to consider its influence when quantizing the signal, that is, one thus

may want to evaluate the difference $\hat{d}(n) - d(n)$ also beyond the current frame, $n \geq K + N$. However, to do that, one mey want to consider the influence of the residual of the current frame only by setting residuals of the following frames to zero. Therefore, the ZIR of $\hat{d}(n)$ into the next frame may be compared. In other words, the modified target is obtained:

$$d'(n) = \begin{cases} 0 & n < K \\ d(n) & K \leq n < K + N \\ -\sum_{k=1}^{m} a(k) d'(n-k) & n > K + N. \end{cases} \quad \text{(10d)}$$

Equivalently, using the impulse response h(n) of A(z), then

$$d'(n) = \sum_{k=K}^{K+N-1} e(k) h(n-k). \quad \text{(10e)}$$

This formula can be written in a convenient matrix form by $d' = He$ where H and e are defined as in Equation 2. It can be seen that the modified target is exactly x of Equation 2.

In calculation of matrix R, note that in theory, the impulse response h(k) is an infinite sequence, which is not realisable in a practical system.

However, either

1) truncating or windowing the impulse response to a finite length and determining the autocorrelation of the truncated impulse response, or

2) calculating the power spectrum of the impulse response using the Fourier spectra of the associated LP and perceptual filters, and obtain the autocorrelation by an inverse Fourier transform

is possible.

Now, an extension employing LTP is described.

The long-time predictor (LTP) is actually also a linear predictor.

According to an embodiment, the matrix determiner 110 may be configured to determine the autocorrelation matrix *R* depending on a perceptually weighted linear predictor, for example, depending on the long-time predictor.

The LP and LTP can be convolved into one joint predictor, which includes both the spectral envelope shape as well as the harmonic structure. The impulse response of such a predictor will be very long, whereby it is even more difficult to handle with prior art. However, if the autocorrelation of the linear predictor is already known, then the autocorrelation of the joint predictor can be calculated by simply filtering the autocorrelation with the LTP forward and backward, or with a similar process in the frequency domain.

Note that prior methods employing LTP have a problem when the LTP lag is shorter than the frame length, since the LTP would cause a feedback loop within the frame. The benefit of including the LTP in the objective function is that when the lag of the LTP is shorter than frame length, then this feedback is explicitly taken into account in the optimisation.

In the following, an extension for fast optimisation in an uncorrelated domain is described.

A central challenge in design of ACELP systems has been reduction of computational complexity. ACELP systems are complex because filtering by LP causes complicated correlations between the residual samples, which are described by the matrix B or in the current context by matrix *R*. Since the samples of e(n) are correlated, it is not possible to just quantise e(n) with desired accuracy, but many combinations of different quantisations with a trial-and-error approach have to be tried, to find the best quantisation with respect to the objective function of Equation 3 or 10, respectively.

By the introduction of the matrix *R*, a new perspective to these correlations is obtained. Namely, since *R* has Hermitian Toeplitz structure, several efficient matrix decompositions can be applied, such as the singular value decomposition, Cholesky decomposition or Vandermonde decomposition of Hankel matrices (Hankel matrices are upside-down Toeplitz matrices, whereby the same decompositions can be applied to Toeplitz and Hankel matrices) (see [6] and [7]). Let $R = E\,D\,E^H$ be a decomposition of R such that D is a diagonal matrix of the same size and rank as R. Equation 9 can then be modified as follows:

$$\epsilon^2 = (e-\hat{e})^H R (e-\hat{e}) = (e-\hat{e})^H E D E^H (e-\hat{e}) = (f-\hat{f}) D (f-\hat{f}) \qquad (11)$$

where $\hat{f} = E^H \hat{e}$. Since D is diagonal, the error for each sample of f(k) is independent of other samples f(i). In Equation 10, it is assumed that the codebook vector is scaled by the optimal gain, whereby the new objective function is

$$\frac{(f^H D \hat{f})^2}{\hat{f}^H D \hat{f}}. \qquad (12)$$

Here, the samples are again correlated (since changing the quantization of one line changes the optimal gain for all lines), but in comparison to Equation 10, the effect of correlation is here limited. However, even if the correlation is taken into account, optimisation of this objective function is much simpler than optimisation of Equations 3 or 10.

Using this decomposition approach, it is possible

1. to apply any conventional scalar or vector quantization technique with desired accuracy, or
2. to use Equation 12 as the objective function with any conventional ACELP pulse search algorithm.

Both approaches give a near-optimal quantization with respect to Equation 12. Since conventional quantization techniques generally do not require any brute-force methods (for the exception of a possible rate-loop), and because the matrix D is simpler than either B or R, both quantization methods are less complex than conventional ACELP pulse search algorithms. The main source of computational complexity in this approach is thus the computation of the matrix decomposition.

Some embodiments employ equation 12 to determine a codebook vector of the codebook.

E.g., several matrix factorizations for R of the form $R = E^H D E$ exist. For example,

(a) The eigenvalue decomposition can be calculated for example by using the GNU Scientific Library (http://www.gnu.org/software/gsl/manual/html_node/Real-Symmetric-Matrices.html). The matrix R is real and symmetric (as well as Toeplitz), whereby the function “gsl_eigen_symm()” can be used to determine the

matrices E and D. Other implementations of the same eigenvalue decomposition are readily available in literature [6].

(b) The Vandermonde factorization of Toeplitz matrices [7] can be used using the algorithm described in [8]. This algorithm returns matrices E and D such that E is a Vandermonde matrix, which is equivalent to a discrete Fourier transform with non-uniform frequency distribution.

Using such factorizations, the residual vector e can be transformed to the transform domain by $f = E^H e$ or $f' = D^{1/2} E^H e$. Any common quantization method can be applied in this domains, for example,

1. The vector f' can be quantized by an algebraic codebook exactly as in common implementations of ACELP. However, since the elements of f' are uncorrelated, a complicated search function as in ACELP is not needed, but a simple algorithm can be applied, such as

(a) Set initial gain to g=1

(b) Quantize f' by $\hat{f}' = round(gf')$.

(c) If the number of pulses in f' is larger than a pre-defined amount p, $\|\hat{f}'\|_1 > p$, then increase gain g and return to step b.

(d) Otherwise, if the number of pulses in $\hat{f}'$ is smaller than a pre-defined amount p, $\|\hat{f}'\|_1 < p$, then decrease gain g and return to step b.

(e) Otherwise, the number of pulses in $\hat{f}'$ is equal to the pre-defined amount p, $\|\hat{f}'\|_1 = p$, and processing can be stopped.

2. An arithmetic coder can be used similar to that used in quantization of spectral lines in TCX in the standards AMR-WB+ or MPEG USAC.

It should be noted that since the elements of f' are orthogonal (as can be seen from Equation 12) and they have the same weight in the objective function of Equation 12, they can be quantized separately, and with the same quantization step size. That quantization will automatically find the optimal (the largest) value of the objective function in Equation 12, which is possible with that quantization accuracy. In other words, the quantization

algorithms presented above, will both return the optimal quantization with respect to Equation 12.

This advantage of optimality is tied to the fact that the elements of f' can be treated separately. If a codebook approach would be used, where the codebook vectors c_k are non-trivial (have more than one non-zero elements), then these codebook vectors would not have independent elements anymore and the advantage of the matrix factorization is lost.

Observe that the Vandermonde factorization of a Toeplitz matrix can be chosen such that the Vandermonde matrix is a Fourier transform matrix but with unevenly distributed frequencies. In other words, the Vandermonde matrix corresponds to a frequency-warped Fourier transform. It follows that in this case the vector f corresponds to a frequency domain representation of the residual signal on a warped frequency scale (see the "root-exchange property" in [8]).

Importantly, notice that this consequence is not well-known. In practice, this result states that if a signal x is filtered with a convolution matrix C, then

$$\|C\,x\|^2 = \|D\,V\,x\|^2 \qquad (13)$$

where V is a (e.g., warped) Fourier transform (which is a Vandermonde matrix with elements on the unit circle) and D a diagonal matrix. That is, if it is desired to measure the energy of a filtered signal, the energy of frequency-warped signal can equivalently be measured. In converse, any evaluation that shall be done in a warped Fourier domain, can equivalently be done in a filtered time-domain. Due to the duality of time and frequency, an equivalence between time-domain windowing and time-warping also exists. A practical issue is, however, that finding a convolution matrix C which satisfies the above relationship is a numerically sensitive problem, whereby often it is easier to find approximate solutions $\hat{C}$ instead.

The relation $\|C\ x\|^2 = \|D\ V\ x\|^2$ can be employed for determining a codebook vector of a codebook.

For this, it should first be noted that here, by H, a convolution matrix like in Equation 2 will be denoted instead of C. If, then, one wants to minimize the quantization noise $e = Hx - H\hat{x}$, its energy can be measured:

$$\varepsilon^2 = \|Hx - H\hat{x}\|^2 = \|H(x-\hat{x})\|^2 = (x-\hat{x})^T H^T H(x-\hat{x}) = (x-\hat{x})^T R(x-\hat{x}) = (x-\hat{x})^T V^H DV(x-\hat{x}) = \left\|D^{1/2}V(x-\hat{x})\right\|^2 = \left\|D^{1/2}V(x-\hat{x})\right\|^2 = \left\|D^{1/2}(f-\hat{f})\right\|^2 = \left\|f'-\hat{f}'\right\|^2. \quad (13a)$$

Now, an extension for frame-independence is described.

When the encoded speech signal is transmitted over imperfect transmission lines such as radio-waves, invariably, packets of data will sometimes be lost. If frames are dependent on each other, such that packet N is needed to perfectly decode N-1, then the loss of packet N-1 will corrupt the synthesis of both packets N-1 and N. If, on the other hand, frames are independent, then the loss of packet N-1 will corrupt the synthesis of packet N-1 only. It is therefore important to device methods that are free from inter-frame dependencies.

In conventional ACELP systems, the main source of inter-frame dependency is the LTP and to some extent also the LP. Specifically, since both are infinite impulse response (IIR) filters, a corrupted frame will cause an "infinite" tail of corrupted samples. In practice, that tail can be several frames long, which is perceptually annoying.

Using the framework of the current invention, the path through which inter-frame dependency is generated can be quantified by the ZIR from the current frame into the next is realized. To avoid this inter-frame dependency, three modifications to the conventional ACELP need to be made.

1. When calculating the ZIR from the previous frame into the current (sub)frame, it should be calculated from the original (not quantized) residual extended with zeros, not from the quantized residual. In this way, the quantization errors from the previous (sub)frame will not propagate into the current (sub)frame.

2. When quantizing the current frame, the error in the ZIR into the next frame between the original and quantized signals must be taken into account. This can be done by replacing the correlation matrix B with the autocorrelation matrix R, as explained above. This ensures that the error in the ZIR into the next frame is minimised together with the error within the current frame.

3. Since the error propagation is due to both the LP and the LTP, both components must be included in the ZIR. This is in difference to the conventional approach where the ZIR is calculated for the LP only.

If quantization errors of previous frame when quantizing the current frame are not taken into account, efficiency in perceptual quality of the output is lost. Therefore, it is possible to choose to take previous errors into account when there is no risk of error propagation. For example, conventional ACELP system apply a framing where every 20ms frame is sub-divided into 4 or 5 subframes. The LTP and the residual are quantized and coded separately for each subframe, but the whole frame is transmitted as one block of data. Therefore, individual subframes cannot be lost, but only complete frames. It follows that it is required to use frame-independent ZIRs only at frame borders, but ZIRs can be used with interframe dependencies between the remaining subframes.

Embodiments modify conventional ACELP algorithms by inclusion of the effect of the impulse response of the current frame into the next frame, into the objective function of the current frame. In the objective function of the optimisation problem, this modification corresponds to replacing a correlation matrix with an autocorrelation matrix that has Hermitian Toeplitz structure. This modification has the following benefits:

1. Computational complexity and memory requirements are reduced due to the added Hermitian Toeplitz structure of the autocorrelation matrix.

2. The same perceptual model will be applied on all samples, making the design and tuning of the perceptual model simpler, and its application more efficient and consistent.

3. Inter-frame correlations can be avoided completely in the quantization of the current frame, by taking into account only the unquantized impulse response from the previous frame and the quantized impulse response into the next frame. This improves robustness of systems where packet-loss is expected.

Fig. 2 illustrates a decoder 220 for decoding an encoded speech signal being encoded by an apparatus according to the above-described embodiment to obtain a decoded speech signal. The decoder 220 is configured to receive the encoded speech signal, wherein the encoded speech signal comprises the an indication of the codebook vector, being determined by an apparatus for encoding a speech signal according to one of the above-described embodiments, for example, an index of the determined codebook vector. Furthermore, the decoder 220 is configured to decode the encoded speech signal to obtain a decoded speech signal depending on the codebook vector.

Fig. 3 illustrates a system according to an embodiment. The system comprises an apparatus 210 according to one of the above-described embodiments for encoding an input speech signal to obtain an encoded speech signal. The encoded speech signal comprises an indication of the determined codebook vector determined by the apparatus 210 for encoding a speech signal, e.g., it comprises an index of the codebook vector. Moreover, the system comprises a decoder 220 according to the above-described embodiment for decoding the encoded speech signal to obtain a decoded speech signal. The decoder 220 is configured to receive the encoded speech signal. Moreover, the decoder 220 is configured to decode the encoded speech signal to obtain a decoded speech signal depending on the determined codebook vector.

Although some aspects have been described in the context of an apparatus, these aspects also represent a description of the corresponding method, where a block or device corresponds to a method step or a feature of a method step. Analogously, aspects described in the context of a method step also represent a description of a corresponding block or item or feature of a corresponding apparatus.

The inventive decomposed signal can be stored on a digital storage medium or can be transmitted on a transmission medium such as a wireless transmission medium or a wired transmission medium such as the Internet.

Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a digital storage medium, for example a floppy disk, a DVD, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed.

Some embodiments according to the invention comprise a non-transitory data carrier having electronically readable control signals, which are capable of cooperating with a programmable computer system, such that one of the methods described herein is performed.

Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing one of the methods when the computer program product runs on a computer. The program code may for example be stored on a machine readable carrier.

Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine readable carrier.

In other words, an embodiment of the inventive method is, therefore, a computer program having a program code for performing one of the methods described herein, when the computer program runs on a computer.

A further embodiment of the inventive methods is, therefore, a data carrier (or a digital storage medium, or a computer-readable medium) comprising, recorded thereon, the computer program for performing one of the methods described herein.

A further embodiment of the inventive method is, therefore, a data stream or a sequence of signals representing the computer program for performing one of the methods described herein. The data stream or the sequence of signals may for example be configured to be transferred via a data communication connection, for example via the Internet.

A further embodiment comprises a processing means, for example a computer, or a programmable logic device, configured to or adapted to perform one of the methods described herein.

A further embodiment comprises a computer having installed thereon the computer program for performing one of the methods described herein.

In some embodiments, a programmable logic device (for example a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

The above described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

References

[1] Salami, R. and Laflamme, C. and Bessette, B. and Adoul, J.P., "ITU-T G. 729 Annex A: reduced complexity 8 kb/s CS-ACELP codec for digital simultaneous voice and data", Communications Magazine, IEEE, vol 35, no 9, pp 56-63, 1997.

[2] 3GPP TS 26.190 V7.0.0 , "Adaptive Multi-Rate (AMR-WB) speech codec", 2007.

[3] ITU-T G.718, "Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s", 2008.

[4] Schroeder, M. and Atal, B., "Code-excited linear prediction (CELP): High-quality speech at very low bit rates", Acoustics, Speech, and Signal Processing, IEEE Int Conf, pp 937-940, 1985.

[5] Byun, K.J. and Jung, H.B. and Hahn, M. and Kim, K.S., "A fast ACELP codebook search method", Signal Processing, 2002 6th International Conference on, vol 1, pp 422-425, 2002.

[6] G. H. Golub and C. F. van Loan, "Matrix Computations", 3rd Edition, John Hopkins University Press, 1996.

[7] Boley, D.L. and Luk, F.T. and Vandevoorde, D., "Vandermonde factorization of a Hankel matrix", Scientific computing, pp 27-39, 1997.

[8] Bäckström, T. and Magi, C., "Properties of line spectrum pair polynomials - A review", Signal processing, vol. 86, no. 11, pp. 3286-3298, 2006.

[9] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. Laine, and J. Huopaniemi, "Frequencywarped signal processing for audio applications," J. Audio Eng. Soc, vol. 48, no. 11, pp. 1011–1031, 2000.

[10] T. Laakso, V. Välimäki, M. Karjalainen, and U. Laine, "Splitting the unit delay [FIR/all pass filters design]," IEEE Signal Process. Mag.,vol. 13, no. 1, pp. 30–60, 1996.

[11] J. Smith III and J. Abel, "Bark and ERB bilinear transforms," IEEE Trans. Speech Audio Process., vol. 7, no. 6, pp. 697–708, 1999.

[12] R. Schappelle, "The inverse of the confluent Vandermonde matrix," IEEE Trans. Autom. Control, vol. 17, no. 5, pp. 724–725, 1972.

[13] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," Speech and Audio Processing, IEEE Transactions on, vol. 10, no. 8, pp. 620–636, 2002.

[14] M. Bosi and R. E. Goldberg, Introduction to Digital Audio Coding and Standards. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2003.

[15] B. Edler, S. Disch, S. Bayer, G. Fuchs, and R. Geiger, "A time-warped MDCT approach to speech transform coding," in Proc 126th AES Convention, Munich, Germany, May 2009.

[16] J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE, vol. 63, no. 4, pp. 561–580, April 1975.

[17] J.-P. Adoul, P. Mabilleau, M. Delprat, and S. Morissette, "Fast CELP coding based on algebraic codes," in Acoustics, Speech, and Signal Processing, IEEE Int Conf (ICASSP'87), April 1987, pp. 1957–1960.

[18] ISO/IEC 23003-3:2012, "MPEG-D (MPEG audio technologies), Part 3: Unified speech and audio coding," 2012.

[19] F.-K. Chen and J.-F. Yang, "Maximum-take-precedence ACELP: a low complexity search method," in Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on, vol. 2. IEEE, 2001, pp. 693–696.

[20] R. P. Kumar, "High computational performance in code exited linear prediction speech model using faster codebook search techniques," in Proceedings of the International Conference on Computing: Theory and Applications. IEEE Computer Society, 2007, pp. 458–462.

[21] N. K. Ha, "A fast search method of algebraic codebook by reordering search sequence," in Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on, vol. 1. IEEE, 1999, pp. 21–24.

[22] M. A. Ramirez and M. Gerken, "Efficient algebraic multipulse search," in Telecommunications Symposium, 1998. ITS'98 Proceedings. SBT/IEEE International. IEEE, 1998, pp. 231–236.

[23] ITU-T Recommendation G.191, "Software tool library 2009 user's manual," 2009.

[24] ITU-T Recommendation P.863, "Perceptual objective listening quality assessment," 2011.

[25] T. Thiede, W. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandeburg et al., "PEAQ – the ITU standard for objective measurement of perceived audio quality," Journal of the Audio Engineering Society, vol. 48, 2012.

[26] ITU-R Recommendation BS.1534-1, "Method for the subjective assessment of intermediate quality level of coding systems," 2003.

CLAIMS:

1. An apparatus for encoding a speech signal by determining a codebook vector of a speech coding algorithm, wherein the apparatus comprises:

a matrix determiner for determining an autocorrelation matrix R, and

a codebook vector determiner for determining the codebook vector depending on the autocorrelation matrix R,

wherein the apparatus is configured to encode the speech signal by generating an encoded speech signal, such that the encoded speech signal comprises a plurality of Linear Prediction coefficients, an indication of a fundamental frequency of voiced sounds, and an indication of said codebook vector, being determined by the codebook vector determiner,

wherein the matrix determiner is configured to determine the autocorrelation matrix R by determining vector coefficients of a vector r, wherein the autocorrelation matrix R comprises a plurality of rows and a plurality of columns, wherein the vector r indicates one of the columns or one of the rows of the autocorrelation matrix R, wherein

$$R(i, j) = r(|i-j|),$$

wherein $R(i, j)$ indicates the coefficients of the autocorrelation matrix R, wherein i is a first index indicating one of a plurality of rows of the autocorrelation matrix R, and wherein j is a second index indicating one of the plurality of columns of the autocorrelation matrix R,

wherein the codebook vector determiner is configured to determine the codebook vector by applying the formula

$$f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$$

wherein R is the autocorrelation matrix, and wherein $\hat{e}$ is one of the codebook vectors of the speech coding algorithm, wherein $f(\hat{e})$ is a normalized correlation, and wherein d^T is defined according to

$$\frac{(e^T R \hat{e})^2}{(\hat{e}^T R \hat{e})} = \frac{(d^T e)^2}{(\hat{e}^T R \hat{e})},$$

wherein e is an original, unquantized residual signal.

2. An apparatus according to claim 1,

wherein the codebook vector determiner is configured to determine that codebook vector $\hat{e}$ of the speech coding algorithm which maximizes the normalized correlation

$$f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$$

3. An apparatus according to claim 1 or claim 2,

wherein the matrix determiner is configured to determine the vector coefficients of the vector r by applying the formula:

$$r(k) = h(k) \cdot h(-k) = \sum_i h(i)h(i-k)$$

wherein h(k) indicates a perceptually weighted impulse response of a linear predictive model, and wherein k is an index being an integer, and wherein l is an index being an integer.

4. An apparatus according to any one of claims 1 to 3,

wherein the matrix determiner is configured to determine the autocorrelation matrix R depending on a perceptually weighted linear predictor.

5. An apparatus according to any one of claims 1 to 4, wherein the codebook vector determiner is configured to decompose the autocorrelation matrix R by conducting a matrix decomposition.

6. An apparatus according to any one of claims 1 to 5, wherein the codebook vector determiner is configured to determine the codebook vector depending on a zero impulse response of the speech signal.

7. An apparatus according to any one of claims 1 to 6,

wherein the apparatus is an encoder for encoding the speech signal by employing algebraic code excited linear prediction speech coding, and

wherein the codebook vector determiner is configured to determine the codebook vector based on the autocorrelation matrix R as a codebook vector of an algebraic codebook.

8. A system, comprising:

an apparatus according to any one of claims 1 to 7 for encoding an input speech signal to obtain an encoded speech signal, and

a decoder for decoding the encoded speech signal to obtain a decoded speech signal,

wherein the decoder is configured to receive the encoded speech signal, wherein the encoded speech signal comprises an indication of a codebook vector, being determined by the apparatus according to any one of claims 1 to 7,

wherein the decoder is configured to decode the encoded speech signal to obtain the decoded speech signal depending on the codebook vector.

9. A method for encoding a speech signal by determining a codebook vector of a speech coding algorithm, wherein the method comprises:

determining an autocorrelation matrix R, and

determining the codebook vector depending on the autocorrelation matrix R,

encoding the speech signal by generating an encoded speech signal, such that the encoded speech signal comprises a plurality of Linear Prediction coefficients, an indication of a fundamental frequency of voiced sounds, and an indication of said codebook vector,

wherein determining an autocorrelation matrix R comprises determining vector coefficients of a vector r, wherein the autocorrelation matrix R comprises a plurality of rows and a plurality of columns, wherein the vector r indicates one of the columns or one of the rows of the autocorrelation matrix R, wherein

$R(i, j) = r(|i-j|)$,

wherein $R(i, j)$ indicates the coefficients of the autocorrelation matrix R, wherein i is a first index indicating one of a plurality of rows of the autocorrelation matrix R, and wherein j is a second index indicating one of the plurality of columns of the autocorrelation matrix R,

wherein determining the codebook vector is conducted by applying the formula

$$f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$$

wherein R is the autocorrelation matrix, and wherein $\hat{e}$ is one of the codebook vectors of the speech coding algorithm, wherein $f(\hat{e})$ is a normalized correlation, and wherein d^T is defined according to

$$\frac{(e^T R \hat{e})^2}{(\hat{e}^T R \hat{e})} = \frac{(d^T e)^2}{(\hat{e}^T R \hat{e})},$$

wherein e is an original, unquantized residual signal.

10. A method comprising:

encoding an input speech signal according to the method of claim 9 to obtain an encoded speech signal, wherein the encoded speech signal comprises an indication of a codebook vector, and

decoding the encoded speech signal to obtain a decoded speech signal depending on the codebook vector.

11. A computer-readable medium storing statements and instructions for use, in the execution in a computer, of the method as claimed in claim 9 or claim 10.

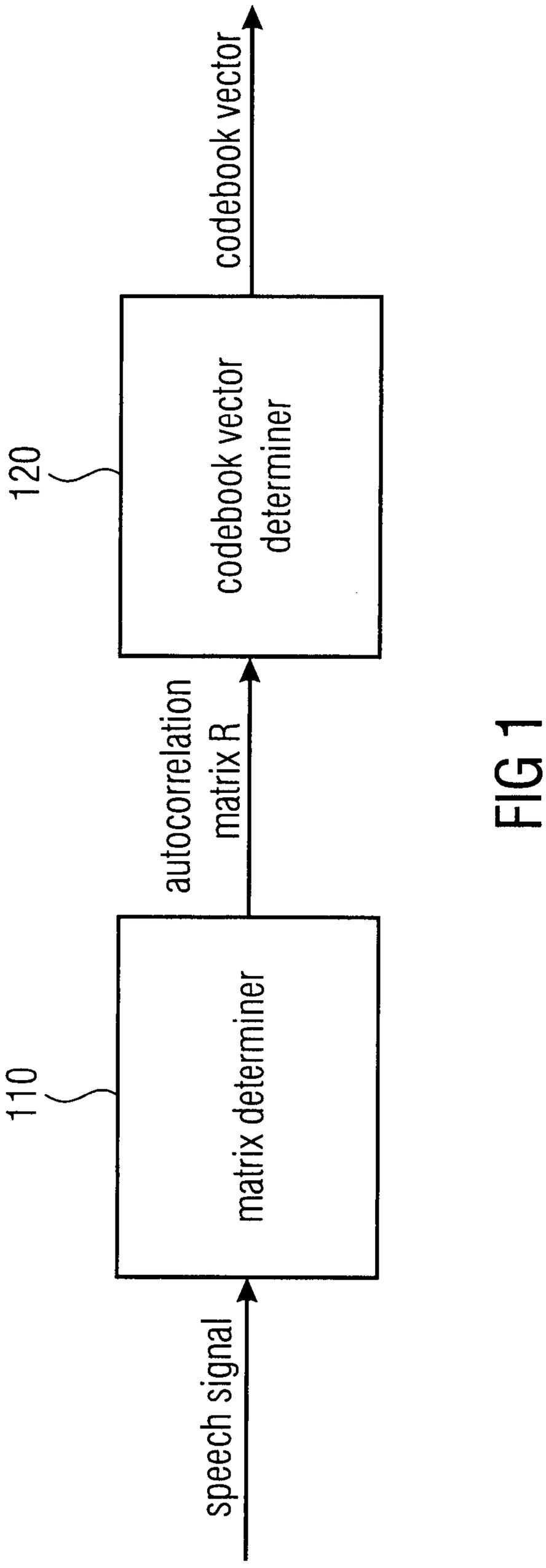

speech signal
110
matrix determiner
autocorrelation matrix R
120
codebook vector determiner
codebook vector

FIG 1

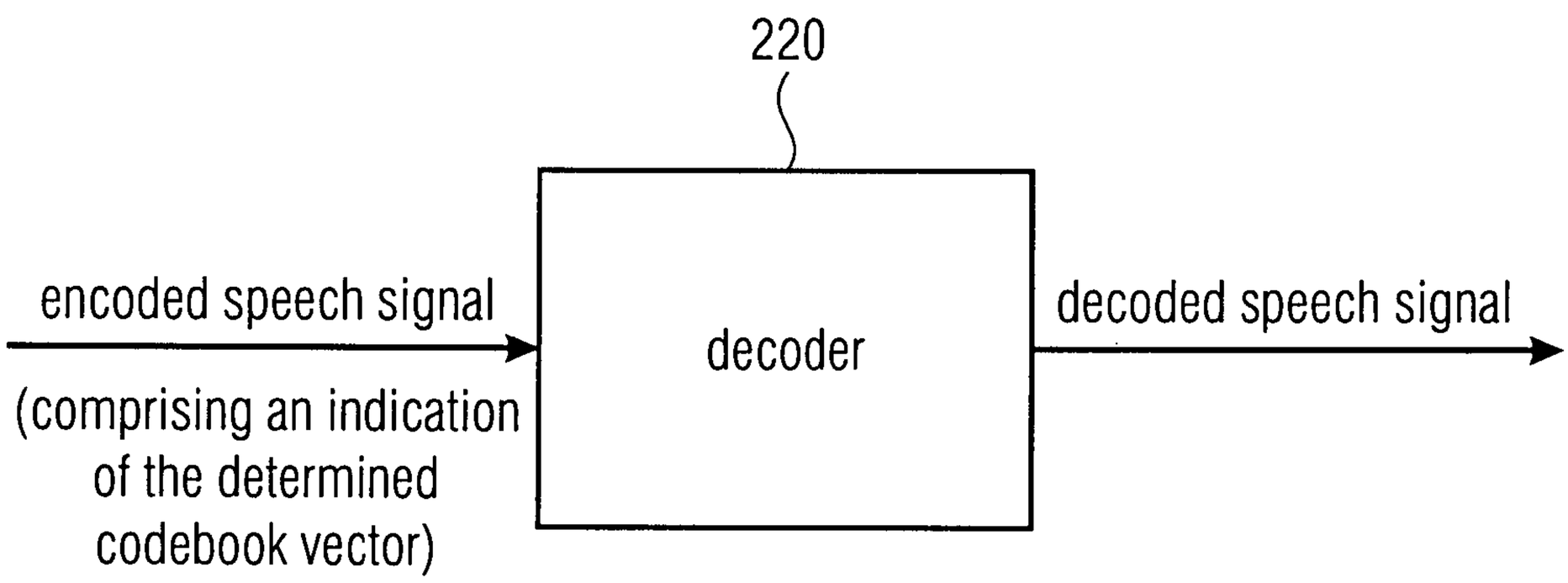

220
encoded speech signal
(comprising an indication of the determined codebook vector)
decoder
decoded speech signal

FIG 2

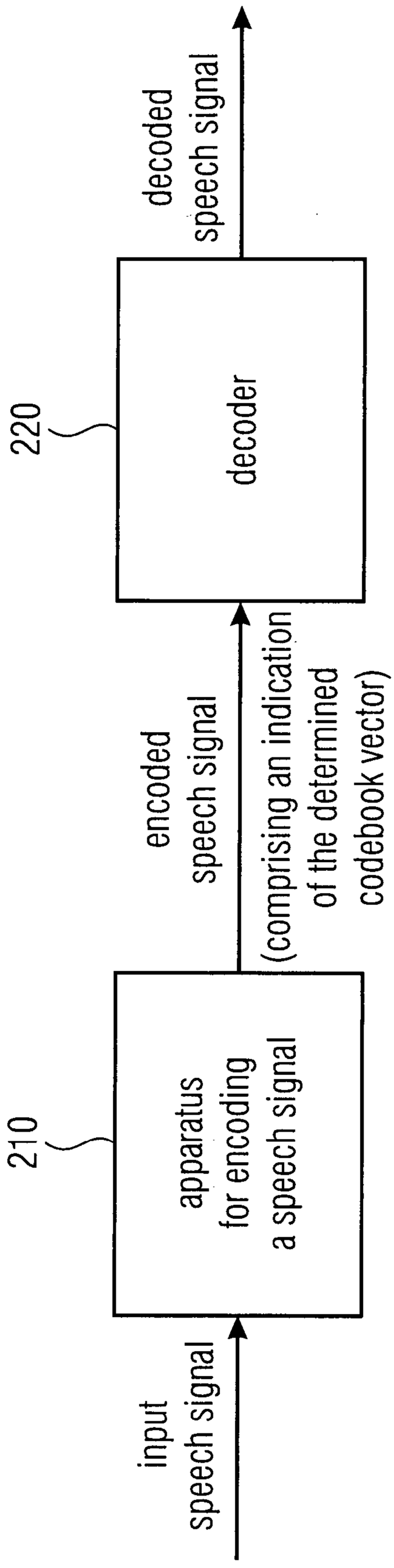
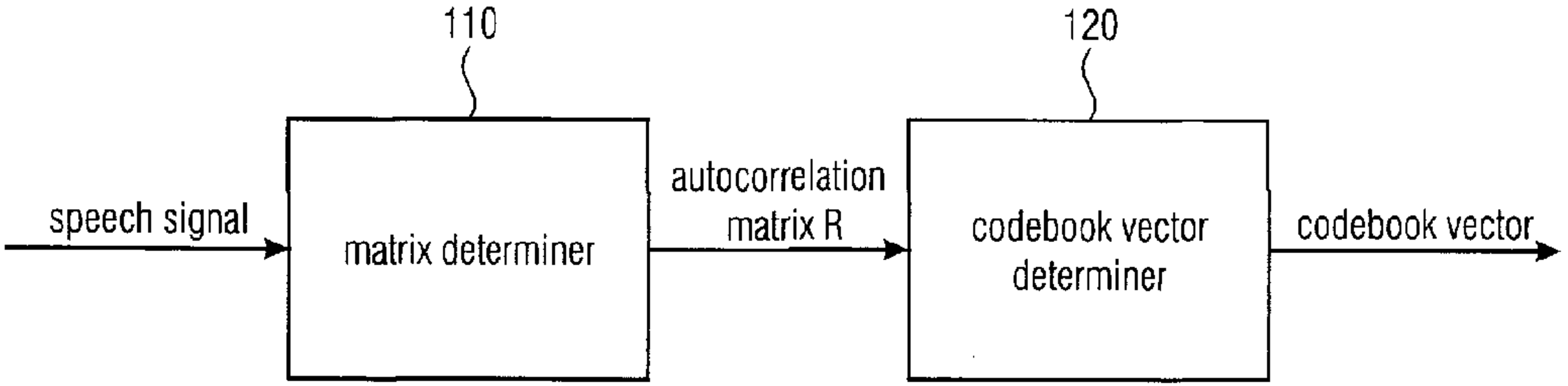

input
speech signal
210
apparatus
for encoding
a speech signal
encoded
speech signal
(comprising an indication
of the determined
codebook vector)
220
decoder
decoded
speech signal

FIG 3


110
120
speech signal
matrix determiner
autocorrelation matrix R
codebook vector determiner
codebook vector