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

(11) **EP 1 970 893 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
17.09.2008 Bulletin 2008/38

(51) Int Cl.:
G10L 11/00 *(2006.01)*

(21) Application number: **07450044.8**

(22) Date of filing: **13.03.2007**

(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

(71) Applicants:
- **Österreichische Akademie der Wissenschaften**
  1010 Wien (AT)
- **Austria Wirtschaftsservice Gesellschaft m.b.H.**
  1030 Wien (AT)

(72) Inventor: **Weruaga, Luis, Dr.**
1090 Wien (AT)

(74) Representative: **Weiser, Andreas**
Patentanwalt,
Hietzinger Hauptstrasse 4
1130 Wien (AT)

(54) **A method for estimating signal coding parameters**

(57) A method for estimating coding parameters of a predictive filter model of a digital signal, in particular speech signal, comprises:
receiving a segment of said signal;
computing the spectrum of said segment;
estimating the background noise in said segment; and
estimating the fundamental frequency in said segment;
computing a spectral mask on the basis of said background noise and said fundamental frequency; and
determining those coding parameters that substantially minimize a cost function which is based on said spectrum, said spectral mask and said predictive filter model.

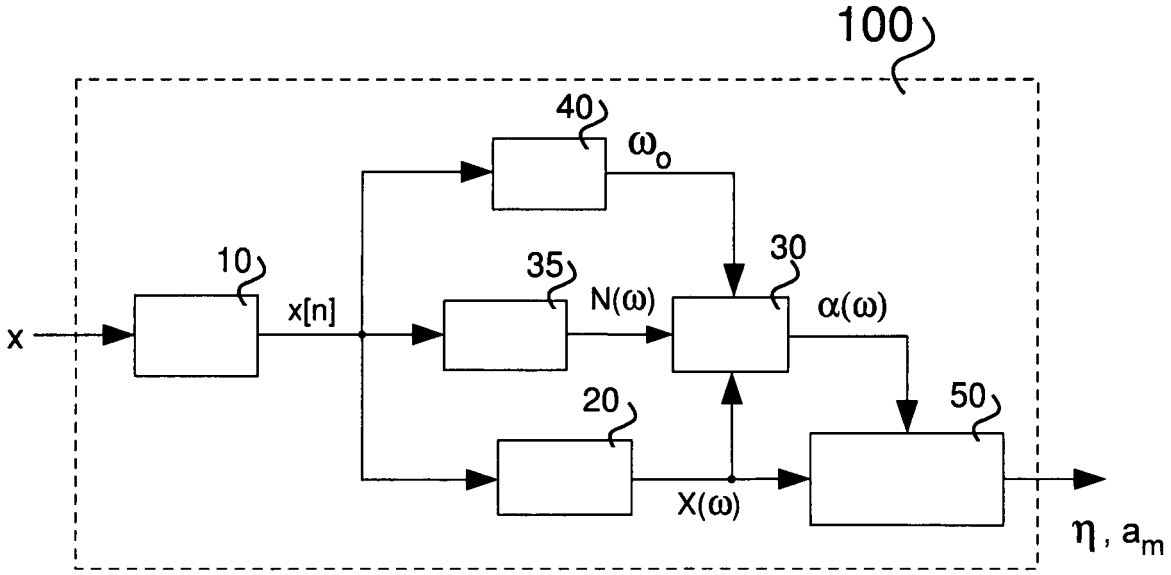

Fig. 1

EP 1 970 893 A1

Description

[0001] The present invention relates to an improved technique for encoding a digital signal, in particular a speech signal. More specifically, the invention concerns a method for estimating coding parameters of a predictive filter model of a digital signal according to the preamble of claim 1.

[0002] A widely-used technique for speech coding is the so-called Linear Predictive Coding (LPC). Said technique computes the parameters of an autoregressive filter from the time samples of a digital speech signal. The computation of those parameters is well-known to those of ordinary skill in the field of the present invention. An example of such computation is found in *ITU-T Recommendation G.722.2,* "Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB)", Geneva 2002. Most of the commercial speech coders such as the LPC Vocoder, the Coded-Excited Linear Predictive Coding (CELP), and its posterior variants (ACELP, VSELP), among many others, rely on the LPC technique.

[0003] In general, the LPC technique is founded on the minimization of the energy of the prediction error e[n]

$$e[n] = x[n] - \sum_{m=1}^{M} a_m x[n-m] \qquad (1)$$

where x[n] is a windowed segment of the input digital signal, a_m are the linear prediction coefficients and M is the model order. In the later decoding, i.e. synthesis stage, the signal is re-generated on the basis of these coefficients input to the synthesis equivalent of the predictive filter model, which synthesis equivalent is defined by the transfer function

$$H(\omega) = \frac{1}{1 - \sum_{m=1}^{M} a_m e^{-j\omega m}} \qquad (2)$$

[0004] The energy of said prediction error can be formulated, by using Parseval's relation, in the frequency domain as a cost function

$$E = \int_{2\pi} \frac{|X(\omega)|^2}{|H(\omega)|^2} d\omega \qquad (3)$$

with $X(\omega)$ being the spectral transformation of the signal segment $x[n]$.

[0005] According to the mentioned equivalence between time and frequency, the solution delivered by the LPC technique is thus equivalent to the linear prediction coefficients that make cost function E minimal.

[0006] Speech coders based on the LPC technique are known to deliver coded speech of acceptable but moderate quality. Furthermore, the performance of automatic speech recognition systems drops notably when fed with said coded signal instead of the raw signal. The author of the present invention found out that although a predictive filter model is adequate for describing the physical production of speech, the LPC technique is unable to obtain the parameters of said model with enough accuracy.

[0007] It is therefore an object of the invention to determine coding parameters for digital signals, in particular speech signals, with improved accuracy.

[0008] This object is achieved by means of a method for estimating coding parameters of a predictive filter model of a digital signal, in particular speech signal, comprising:

receiving a segment of said signal;
computing the spectrum of said segment;
estimating the background noise in said segment; and

estimating the fundamental frequency in said segment;

which method is characterized by the steps of:

computing a spectral mask on the basis of said background noise and said fundamental frequency; and
determining those coding parameters that substantially minimize a cost function which is based on said spectrum, said spectral mask and said predictive filter model.

[0009] In the present disclosure the term "substantially minimizing" is intended to comprise both, making the cost function minimal as well as making the cost function at least a sufficiently low value, i.e. a value within a given or acceptable tolerance interval from that minimum.

[0010] Thus, the proposed coding of each signal segment comprises two main processing steps: on the one hand, the computation of a spectral mask that weights the relevance of each spectral sample of said segment spectrum, wherein the relevance is determined on the basis of the fundamental frequency of the speech utterance and the spectral characteristics of the noise in the segment, and on the other hand the computation of the coding parameters that make a specific cost function minimal, or at least an appropriate level, where said cost function is built with said segment spectrum, said spectral mask and the parametric filter model.

[0011] The invention is based on the insight that not all spectral samples in an input spectrum necessarily contain valuable information for the estimation of linear prediction coefficients: for instance, the spectrum of voiced speech utterances contains only valuable information at harmonic frequencies, and in the presence of background noise the spectrum of the speech can be corrupted at certain frequency components if its level is lower than that of the noise at said components.

[0012] In contrast to conventional LPC techniques which are severely affected by these effects, the novel frequency-selective approach of the invention increases the coding precision and efficiency, especially in the case of voiced utterances and/or of noise-corrupted signal segments. The method of the invention computes the speech coding parameters on the basis of the spectrum of the signal segment, where said parameters are related to the popular speech formation model, with significantly improved accuracy.

[0013] The method of the invention can replace e.g. the LPC technique in those speech/audio coders that operate with said technique. The invention can also be used in speech/audio coders that do not operate with said LPC technique, such as Harmonic Coders and Hybrid Coders.

[0014] Apart therefrom, the improved accuracy of the estimation of the coding parameters also implies a more accurate estimation of the spectral energy. Thus, the present invention can be used in automatic speech recognition systems also in such a way that spectral-like features can be drawn directly from the estimated filter model and gain level instead of from the signal spectrum.

[0015] In a preferred embodiment of the invention said coding parameters are the gain level and the filter coefficients of said predictive filter model.

[0016] In a particular preferred embodiment said cost function is

$$ML = \int_{(2\Pi)} \alpha(\omega)\left(\frac{|X(\omega)|^2}{\eta|H(\omega)|^2} - \log \frac{|X(\omega)|^2}{\eta|H(\omega)|^2} \right) d\omega$$

with

$X(\omega)$ being said spectrum,
$\alpha(\omega)$ being said spectral mask,
η being said gain level, and
$H(\omega)$ being the transfer function, based on said filter coefficients, of the synthesis equivalent of the predictive filter model.

[0017] This new approach of cost function minimization is on the one hand a processing task which is readily feasible with state-of-the-art processing hardware and/or software, and on the other hand ensures the estimation of coding parameters with remarkable improved accuracy.

[0018] According to a further preferred feature of the invention said spectral mask is chosen as

$$\alpha(\omega) = \rho(\omega)\sum \delta(\omega - k\omega_0)$$

with ω_0 being said fundamental frequency and $\rho(\omega)$ being a noise mask based on said background noise. In particular, said noise mask is preferably

$$\rho(\omega) = \left\{ \begin{array}{l} +1\,,\ if\ |X(\omega)|^2 >> N(\omega) \\ 0\,,\ otherwise \end{array} \right\}$$

with $X(\omega)$ being said spectrum and $N(\omega)$ being the power spectrum of said background noise.

[0019] The step of minimizing said cost function can be performed by means of any suitable algorithm of the art; preferably, a multivariate Newton-Raphson algorithm is used.

[0020] In general, the coding parameters determined according to present invention can be translated into any parameterization which are needed for the subsequent decoding, i.e. synthesis stage. Particularly, it is preferred that said predictive filter model is defined by its synthesis equivalent being a parametric all-pole filter model, an autoregressive coefficients filter (ARC) model, a reflection coefficients filter (RC) model, and/or a line spectral frequencies (LSF) model using the coding parameters determined.

[0021] Further details and advantages of the invention will become apparent from the appended claims and the following detailed description of a preferred embodiment under reference to the enclosed drawings in which

Fig. 1 illustrates the analysis stage of a simplified generic speech/audio coder containing the method for computing the parameters of the speech production model in accordance with the present invention, and
Figs. 2a-d show the superior performance obtained with the present invention in an example scenario.

[0022] From the field of bioacoustics it is known that the biological hearing sense responds to the logarithm of the sound intensity. The invention is based on the insight that this bioacoustic principle of logarithmic sense can be introduced into a maximum-likelihood (ML) correspondence according to equations (1) to (3) between the spectral samples $X(\omega)$ and the synthesis part $H(\omega)$ of the prediction filter model, resulting in

$$ML = -\int_{(2\Pi)} \alpha(\omega) \log P[\varepsilon(\omega)] d\omega \qquad (4)$$

where $\varepsilon(\omega)$ is the spectral residue defined as

$$\varepsilon(\omega) = \log\left(|X(\omega)|^2\right) - \log\left(\eta |H(\omega)|^2\right) \qquad (5)$$

$\alpha(\omega)$ being a spectral mask and P[] denoting the probability density function (PDF).

[0023] Given that the spectral samples $X(\omega)$ are commonly characterized by a Gaussian random variable, the PDF of the logarithmic residual is

$$P[\varepsilon] = \exp(\varepsilon - \exp\varepsilon) \qquad (6)$$

[0024] According to the maximum likelihood criterion (4), the ML functional can now be set up as the following cost function:

$$ML = \int_{(2\Pi)} \alpha(\omega)\left(\frac{|X(\omega)|^2}{\eta|H(\omega)|^2} - \log \frac{|X(\omega)|^2}{\eta|H(\omega)|^2} \right) d\omega \qquad (7)$$

[0025] The spectral mask $\alpha(\omega)$ plays a vital role in the cost function ML in that it contains for each frequency a value that weights the relevance of the spectral sample at said frequency.

[0026] The gain level η and the parameters a_m that define the synthesis filter $H(\omega)$ correspond to the parametric degrees of freedom of the cost function ML. As will be apparent for one skilled in the art, any reference in this disclosure to the cost function ML also comprises any mathematically or technically equivalent expression of equation (7), e.g. a cost function differing from equation (7) in an additive term that does not depend on said parametric degrees of freedom.

[0027] Fig. 1 shows in the form of a block diagram an analysis stage 100 of a speech coder that uses the method of the present invention. A signal segmentation block 10 performs the usual segmentation of an input digital signal x into segments, generally denoted by $x[n]$. A spectral transformation block 20 performs the spectral transformation of said segment. Block 20 performs e.g. a Discrete Fourier Transform, Discrete Sinus Transform and/or a Fan-Chirp Transform, among other popular choices.

[0028] A spectral mask block 30 performs the computation of the spectral mask $\alpha(\omega)$. The segment $x[n]$ is assumed to be corrupted by background noise whose spectral characteristics are described by the power spectrum $N(\omega)$. Furthermore said segment may contain a speech utterance of "voiced" nature, with fundamental frequency ω_0 (in case of "unvoiced" speech utterances, the fundamental frequency is considered zero or very low). Therefore, by making use of the frequency-selective properties of cost function ML, the spectral mask is computed by block 30 as

$$\alpha(\omega) = \rho(\omega)\sum \delta(\omega - k\omega_0) \qquad (8)$$

where $\delta(\omega)$ is the "Dirac delta" function, and $\rho(\omega)$ is the noise mask computed as

$$\rho(\omega) = \begin{Bmatrix} +1, & if\ |X(\omega)|^2 >> N(\omega) \\ 0, & otherwise \end{Bmatrix} \qquad (9)$$

[0029] The goal of said spectral mask is two-fold: on the one hand to disable those spectral samples of the segment spectrum that are sensibly corrupted by noise, and on the other hand to discard the spectral samples that do not correspond to harmonic frequencies. Said harmonic frequencies point out to the high-energy spectral peaks that delineate the spectral envelope of the speech utterance.

[0030] The estimation of the power spectrum $N(\omega)$ is carried out by a noise estimation block 35 according to known ad-hoc techniques, such as a Kalman filter estimation, et cet. The estimation of the fundamental frequency ω_0 is carried out by a pitch analysis block 40 according to known ad-hoc methods, e.g. peak detection of the autocorrelation of the segment, et cet.

[0031] A cost function minimization block 50 carries out the computation of the gain level η and parameters a_m as coding parameters of the filter model that make cost function ML minimal, or at least below a predetermined level. This minimization task is a readily feasible computer programming task. A possible choice for the implementation of the minimization task is the multivariate Newton-Raphson algorithm.

[0032] The output parameters of the speech coder analysis stage 100 are the gain level η, the parameters a_m of the predictive filter, and - if desired - the pitch of the excitation ω_0 which can be taken from the output of block 40. Said parameters correspond to the output of the analysis stage of conventional speech coders e.g. relying on the LPC technique. Therefore, the method of the present invention can supersede the LPC technique in said coders.

[0033] Although all processors of the analysis stage 100 operate with time-discrete and frequency-discrete samples, for the sake of clarity the mathematical description of the invention has been given in continuous frequency. One skilled in the art will immediately recognize that this choice does not affect the essence of the present invention.

Fig. 2a-d illustrate the frequency-selective properties of the present invention on a segment of voiced speech:
Fig. 2a shows an exemplary input signal segment x[n] of 200 samples;
Fig. 2b depicts the logarithmic spectrum envelope obtained with conventional LPC (dotted line) vs. the envelope obtained with the method of the invention (solid line);
Fig. 2c shows the prediction error e[n] with the inventive method; and
Fig. 2d the prediction error e[n] with conventional LPC technique.

[0034] It can be clearly seen that the present invention achieves higher accuracy in estimating the coding parameters of a predictive filter model, manifested by a resulting spectral envelope interpolating narrowly, i.e. matching closely, the energy of the harmonics, see Fig. 2b, and a prediction error closer to the actual excitation, see Fig. 2c.

[0035] The present description contained specific information pertaining both to the scientific basis and the implementation of the present invention. One skilled in the art will recognize that the present invention may be implemented in a manner different from that specifically discussed in the present application. The proposed method can e.g. be implemented and realized efficiently in a digital computer.

[0036] The invention is not limited to the preferred embodiments described in detail above but encompasses all variants and modifications thereof which will become apparent for the man skilled in the art from the present disclosure and which fall into the scope of the appended claims.

Claims

1. A method for estimating coding parameters of a predictive filter model of a digital signal, in particular speech signal, comprising:

receiving a segment of said signal;
computing the spectrum of said segment;
estimating the background noise in said segment; and
estimating the fundamental frequency in said segment;

characterized by the steps of:

computing a spectral mask on the basis of said background noise and said fundamental frequency; and
determining those coding parameters that substantially minimize a cost function which is based on said spectrum, said spectral mask and said predictive filter model.

2. The method of claim 1, wherein said coding parameters are the gain level and the filter coefficients of said predictive filter model.

3. The method of claim 2, wherein said cost function is

$$ML = \int_{(2\Pi)} \alpha(\omega)\left(\frac{|X(\omega)|^2}{\eta|H(\omega)|^2} - \log \frac{|X(\omega)|^2}{\eta|H(\omega)|^2} \right) d\omega$$

with

$X(\omega)$ being said spectrum,
$\alpha(\omega)$ being said spectral mask,
η being said gain level, and
$H(\omega)$ being the transfer function, based on said filter coefficients, of the synthesis equivalent of the predictive filter model.

4. The method of any of the claims 1 to 3, wherein said spectral mask is

$$\alpha(\omega) = \rho(\omega)\sum \delta(\omega - k\omega_0)$$

with ω_0 being said fundamental frequency and $\rho(\omega)$ being a noise mask based on said background noise.

5. The method of claim 4, wherein said noise mask is

$$\rho(\omega) = \left\{ \begin{array}{l} +1\,,\ if\ |X(\omega)|^2 >> N(\omega) \\ 0\,,\ otherwise \end{array} \right\}$$

with $X(\omega)$ being said spectrum and $N(\omega)$ being the power spectrum of said background noise.

6. The method of any of the claims 1 to 5, wherein said step of minimizing said cost function is performed by means of a multivariate Newton-Raphson algorithm.

7. The method of any of the claims 1 to 6, wherein said predictive filter model is defined by its synthesis equivalent being a parametric all-pole filter model, an autoregressive coefficients filter (ARC) model, a reflection coefficients filter (RC) model, and/or a line spectral frequencies (LSF) model.

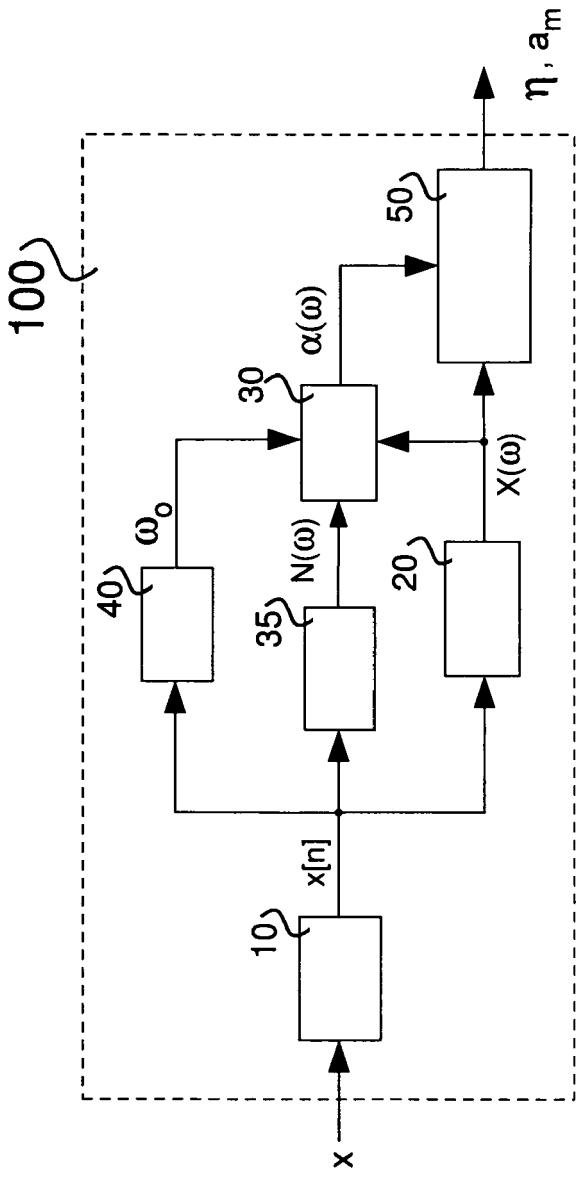

Fig. 1

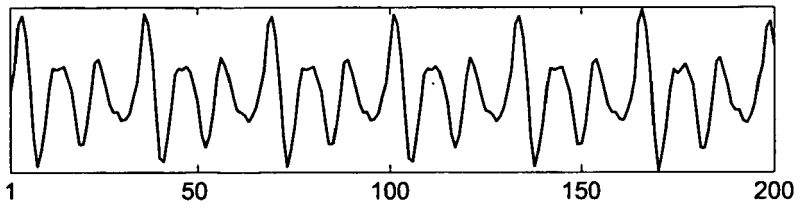

Fig. 2a

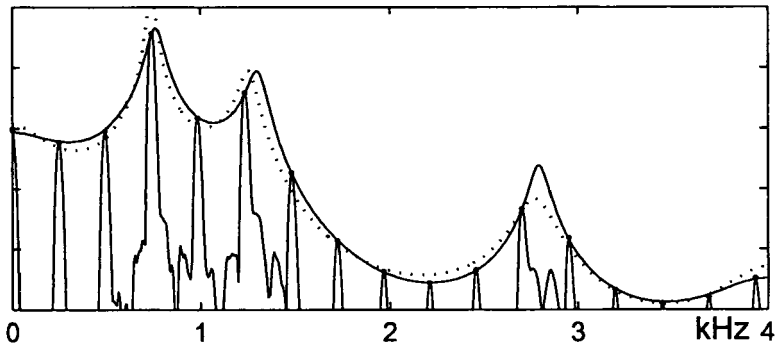

Fig. 2b

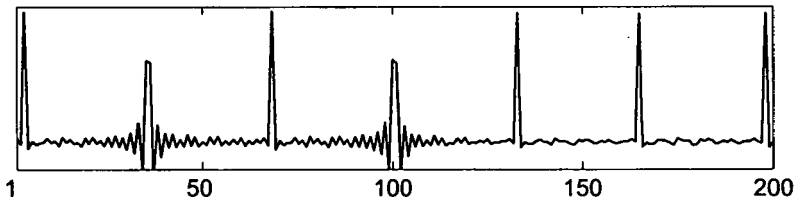

Fig. 2c

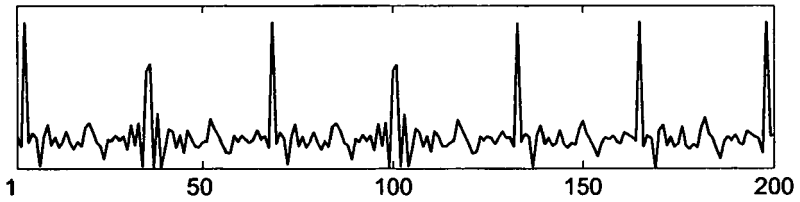

Fig. 2d

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 07 45 0044

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
Y	GU L ET AL: "Perceptual harmonic cepstral coefficients for speech recognition in noisy environment" 2001 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). SALT LAKE CITY, UT, MAY 7 - 11, 2001, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP), NEW YORK, NY : IEEE, US, vol. VOL. 1 OF 6, 7 May 2001 (2001-05-07), pages 125-128, XP010803060 ISBN: 0-7803-7041-4 * page 126, left-hand column, paragraph 5 - page 127, left-hand column, paragraph 2 * * page 127, right-hand column, paragraph 1 - paragraph 4 *	1,2,7	INV. G10L11/00

Y	HERMANSKY H ET AL: "Perceptual linear predictive (PLP) analysis-resynthesis technique" EUROSPEECH 91. 2ND EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY PROCEEDINGS ISTITUTO INT. COMUNICAZIONI GENOVA, ITALY, 1991, pages 329-332 vol.1, XP002442693 * page 1739, left-hand column, paragraph 1 - page 1740, left-hand column, paragraph 2 *	1,2,7	TECHNICAL FIELDS SEARCHED (IPC) G10L
	----- -/--		

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	16 July 2007	Burchett, Stefanie

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)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 07 45 0044

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	LUKASIAK J ET AL: "Linear prediction incorporating simultaneous masking" 2000 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS (CAT. NO.00CH37100) IEEE PISCATAWAY, NJ, USA, vol. 3, 2000, pages 1471-1474 vol., XP002442694 ISBN: 0-7803-6293-4 * page 1472, left-hand column, paragraph 2 - right-hand column, paragraph 4 *	1-7	
A	ZHAO Y: "FREQUENCY-DOMAIN MAXIMUM LIKELIHOOD ESTIMATION FOR AUTOMATIC SPEECH RECOGNITION IN ADDITIVE AND CONVOLUTIVE NOISES" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 8, no. 3, May 2000 (2000-05), pages 255-266, XP011054019 ISSN: 1063-6676 * page 260, right-hand column, paragraph 2 - page 261, left-hand column, paragraph 1 *	1-7	TECHNICAL FIELDS SEARCHED (IPC)
A	EP 0 851 406 A (NIPPON ELECTRIC CO [JP]) 1 July 1998 (1998-07-01) * page 4, line 36 - page 7, line 9 *	1-7	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	16 July 2007	Burchett, Stefanie

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

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 07 45 0044

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.

16-07-2007

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0851406 A	01-07-1998	CA 2225985 A1	27-06-1998
		JP 2914332 B2	28-06-1999
		JP 10190470 A	21-07-1998
		US 6049814 A	11-04-2000

EPO FORM P0459

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