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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 December 2004 (09.12.2004)

PCT

(10) International Publication Number
WO 2004/107318 A1

(51) International Patent Classification[7]: G10L 19/14, 11/02

(21) International Application Number: PCT/IB2003/002336

(22) International Filing Date: 27 May 2003 (27.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **VAN DE PAR, Steven, L., J., D., E.** [NL/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). **SKOWRONEK, Jan, J.** [DE/DE]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: **GROENENDAAL, Antonius, W., M.**; Philips Intellectual Property & Standards, Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

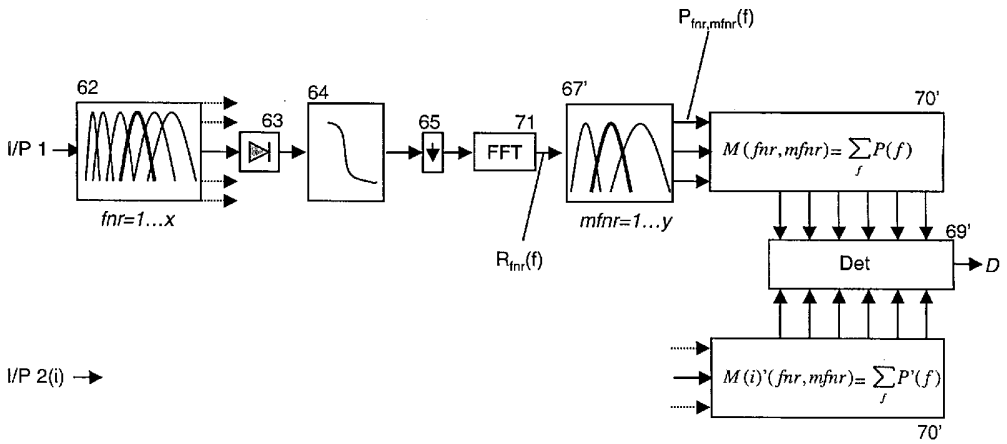
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: AUDIO CODING



(57) Abstract: A method of classifying a spectro-temporal interval of an input audio signal (x(t)) is disclosed. A spectro-temporal interval of the input audio signal is first modelled (62...71) according to a perceptual model to provide a first representation (Rep 1). The spectro-temporal interval is then modelled (62...71) using a modified noise substituted input signal according to the same perceptual model to provide a second representation (Rep 2). The spectro-temporal interval is then classified as being noise or not based on a comparison of the first and second representations.

WO 2004/107318 A1

Audio coding

The present invention relates to a method of coding an audio signal.

The operation of coders such as the MPEG coder is well known. In one implementation, Figure 1, an input PCM (Pulse Code Modulated) signal x(t) is supplied to a sub-band filter bank (SBF) 10 comprising 1024 filters 11 with respective transfer functions $H_1 \ldots H_{1024}$. Each filtered signal is decimated and then supplied to a scaler (SC) 12, which determines appropriate scale factors for each band. Separately, a masking threshold and bit allocation calculator (MT/BA) 13 typically operating with some form of psycho-acoustic model, determines a bit allocation for each frequency band where bit rate is balanced against distortion introduced during quantisation. Each filtered and scaled signal is then quantized (Q) 14 according to the allocated bit rate before being fed to a multiplexer (MUX) 15 where the final audio stream (AS) including quantized signals, scale factors and bit allocation information is generated.

It is known that some spectral and/or temporal parts of audio signals can be represented in a highly efficient manner (e.g. 4 to 10 kb/s) with only a noise model description.

Thus, in relation to Figure 1, the input signal x(t) can be fed to a selection component (Sel) 16 which classifies frequency bands for temporal intervals as either noisy or not. When a spectro-temporal interval is determined to be noisy, the selection component 16 instructs the multiplexer 15 not to code sub-band signals for that interval. The spectro-temporal interval of the input signal x(t) is instead modelled with a noise analyser (NA) 17 whose output is quantized (Q) 18 according to the available bit rate.

A notorious problem, however, is to decide what part of the audio signal can be represented by noise. The decision is based on the assumption that modelling part of the audio signal with noise will not lead to a reduction in quality. In addition, it should also lead to an increase in the efficiency with which the signal can be encoded.

In Schulz, D. "Improving audio codecs by noise substitution", J. Audio Eng. Soc., Vol. 44, pp. 593—598, 1996, it is shown that statistical signal properties of a signal can be derived to make the above classification. Exemplary techniques disclosed by Schulz include:

- Tracking of spectral peaks in successive spectra.
- Using predictors in the frequency domain.
- Using predictability in the time domain with a transversal filter.

In the both the latter examples it is assumed that the more predictable a signal is, the more tonal it is and as such predictability is assumed to be the opposite of noisiness.

Other techniques are based on an analysis of the spectral flatness of a frame (usually over a short duration e.g. 10-20 ms). Again, the flatter the spectrum, the noisier is it considered.

In Herre, J. Schulz, D. "Extending the MPEG-4 AAC codec by perceptual noise substitution", in Proc. 104th convention of the Audio Eng. Soc., Amsterdam, preprint 4720, 1998, the above statistical methods are mentioned in the context of MPEG 4 AAC. Here spectro-temporal intervals correspond to scale-factor-bands and frames and when these are modelled by noise a bit rate saving is made.

It will be seen, however, that the signal statistical criteria of the prior art do not necessarily coincide with criteria that are employed by a human observer i.e. a possible match between these criteria is more or less coincidental.

According to the present invention there is provided a method according to claim 1.

The present invention is based on a noise classification of spectro-temporal intervals of generic audio signals using a perceptual or psycho-acoustical model. The invention is based on predicted audibility of noise substitution, i.e. if noise substitution is predicted to be inaudible to a human observer, it does not lead to perceptual degradation.

Embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 shows a conventional MPEG encoder where selected spectro-temporal portions of an audio signal are represented with noise model parameters;

Figure 2 illustrates the operation of an improved selection component according to an embodiment of the invention operable within the encoder of Figure 1;

Figure 3 is a block diagram of a known psycho-acoustic based signal comparison model;

Figure 4 shows a block diagram of a preferred embodiment of a psycho-acoustic based signal comparison model for use in the selection component of Figure 2.

Figure 5 shows a power spectrum ($R_{fnr}(f)$) of an harmonic tone-complex produced by the FFT component of the model of Figure 4;

Figure 6 shows a power spectrum ($R_{fnr}(f)$) of Gaussian noise produced by the FFT component of the model of Figure 4;

Figure 7 shows an encoder according to a second embodiment of the present invention;

Figure 8 shows the operation of a selection component operable within the encoder of Figure 7; and

Figures 9(a) and 9(b) illustrate the input (R_{25}) and modulation spectrum output ($P_{25,18}$) of one of the filters (25,18) of the filterbank of the model of Figure 4 for an harmonic tone complex and for a noise input signal respectively.

In a first embodiment of the present invention an improved selection component is employed in an MPEG coder of the type shown in Figure 1 to determine whether spectro-temporal intervals can best be modelled through sub-band filtered signals or with a noise model.

Referring now to Figure 2, in general, the improved selection component (Sel) 16' iteratively tests for the substitution of noise modelling for each of a plurality of frequency bands *i* for an interval *n* of input signal x(t). Preferably, the selection component makes its tests over a time period exceeding the basic interval length of the coder.

In the embodiment, an interval t(*n*) of the PCM format input signal x(t) surrounding the test interval *n*, is split into a sequence of 9 short overlapping segments ...s1,s2... These segments are each windowed with a square root Hanning window (or some other analysis window) in segmentation unit 42. (It will be seen that the specific number of intervals is not critical in implementing the invention and for example 8 or 11 intervals could also be used.) At the same time, the signal x(t) for the interval t(*n*) is provided as an input I/P1 to a psycho-acoustic analyser 52.

A FFT (Fast Fourier Transform) is applied on each time-domain windowed signal ...s1,s2..., resulting in respective complex frequency spectrum representations of the windowed signals, step 44.

For each representation and for each frequency band *i*, a noise analyser/synthesizer 46 provides a noise modelled signal for the frequency band *i* with the

remainder of the spectrum unchanged. This noise modelled signal is preferably based on the same model used by the noise analyser (NA) 17 in the encoder proper.

The selection component then takes an inverse FFT of each noise substituted signal to obtain time domain signals ...s'1(*i*),s'2(*i*)..., step 48. In step 50, the separate segments are recombined by first windowing again with a square-root Hanning window (or some other synthesis window) and applying an overlap-add method. This results in a long PCM signal x'(t)(*i*) corresponding to each segment *i* for which noise has been substituted across the interval t(*n*). The signals x'(t)(*i*) are then sent as a series of test input signals I/P2(*i*) to a pyscho-acoustic analyser (PA) 52. In the matrix shown at the lower part of figure 2, a symbolic representation of the modified signal is shown where noise is substituted in the i-th frequency band. Along the horizontal axis, time is depicted, along the vertical axis, the frequency band number (fbnr) corresponding to the scale factor bands used in the AAC encoder. Dots denote areas that contain the original signal samples, the bars depict areas with noise substituted. The grey bar denotes the area to which the noise classification applies.

Within the analyser 52, a perceptual or psycho-acoustic model is used to compute a difference (reduction in quality) between the modified input signals (I/P2(*i*)) and the original signal (I/P1). If this perceptual difference does not exceed a certain criterion value, it is assumed that the middle spectro-temporal interval out of the 9 intervals that have been substituted with noise i.e. the frequency band *i* for interval *n,* can indeed be replaced by noise model parameters. In this fashion all spectro-temporal intervals are studied one by one to make a decision about noise substitutions for all intervals.

It has been found that using the above embodiment where, based on the outcome of the perceptual model, a decision is made for only one of 9 subsituted intervals, a critically more reliable decision about noise substitution is made than by testing and substituting only a single interval at a time.

After all spectro-temporal intervals had been evaluated in this way, the analyser 52 indicates to the multiplexer (MUX), Figure 1, for which of the frequency bands of interval *n* actual noise substitution can be made.

It should be noted that in the preferred embodiment, testing is always performed on the original signal with noise only being substituted in the frequency band *i* being tested, i.e. even if the analyser 52 had determined that noise could be substituted for band *i-1*in interval *n-1*, the original signal would be employed when testing band *i* in interval *n*.

The multiplexer then picks the data to be encoded from either the quantiser 18 for noise analyser NA or the quantiser(s) 14 for the sub-band filter(s) 11 as appropriate and especially with regard to savings in bitrate which may be provided by switching between noise and sub-band filter models.

It will also be seen that the selection component 16' could also be in communication with either or both of the sub-band filters 11 and the noise analyser 17 or the quantisers 14, 18 switching these in and out as appropriate to reduce the overall processing performed by the system. However, this would require the selection component to run ahead of the noise analyser 17 and sub-band filter 10 components and may introduce an undesirable lag in the encoder. Thus, in implementing the embodiment described above lag needs to be balanced against processing overhead.

In a particularly preferred implementation of the first embodiment described above, the perceptual model employed in the analyser 52 is based on a model generally of the type disclosed in Dau, T., Puschel, D., Kohlrausch, A. "A quantitative model of the "effective" signal processing in the auditory system", J. Acoust. Soc. Am., Vol.99, 3615—3631, June 1996; and Dau, T., Kollmeier B., Kohlrausch, A. "Modelling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers", J. Acoust. Soc. Am., Vol.102, 2892—2905, November 1997, Figure 3.

In Dau, an input signal (I/P1 or I/P2) is first sent through an auditory filterbank 62. It is known, that each location on the basilar-membrane inside the human cochlea has a specific bandpass-filter characteristic. The filterbank 62 thus models the frequency-place transformation of the basilar-membrane by producing a plurality x of band-pass filtered time-domain signals which are fed to the next stage in the model. (Each of the next stages in Figure 3 operates on each of the filterbank output signals, however, the processing for only 1 of the x signals is illustrated.)

The next step is a haircell model, comprising half-wave rectification 63, low-pass filtering 64 with a cut-off frequency of 1 kHz and down sampling 65 of each filtered signal. Here the transformation of the mechanical oscillations of the basilar-membrane into receptor potentials in the inner haircells is approximated. The next phase comprises feedback loops 66 to account for the adaptive properties of the auditory periphery.

A modulation or linear filterbank 67 then accounts for the temporal pattern processing of the auditory system. The modulation filterbank comprises a total of y filters divided into two sets, each with different scaling. The first set comprises a filter with a band-width of 2.5 Hz with the next filters going up to 10 Hz having a constant bandwidth of 5 Hz.

The second set, for frequencies between 10 and about 1000 Hz, has a logarithmic scaling where the ratio Q = center frequency/bandwidth=2 is constant, to bring the total to y filters.

In Dau, the modulation filterbank 67 provides a time-domain modulation spectrum. Thus a matrix of $x*y$ of such modulation spectra is produced to represent each input signal. Internal noise 68 is then added to each modulation spectrum signal to model the limited performance resolution of the auditory system.

For each input signal, each matrix representation (Rep 1 and Rep 2) 70 is then fed to a detector 69 which determines the difference (D) between both representations. This quantity can be compared to a pre-determined threshold to indicate whether the difference between signals is audible.

Thus, each individual matrix cell in Dau is a time signal i.e. for each auditory filter and each subsequent modulation filter, there is a time signal resulting from I/P 1 that is compared with a template resulting from I/P 2 to determine whether a certain test-signal (or distortion) is audible.

Thus, if applying Dau straightforwardly to the problem of determining whether noise substitution may be audible, the full temporal structure of a signal would be used in the decision process. Thus, every detail of a substituted noise token could lead to predicted distortion. In reality, listeners are not sensitive to the specific details of a noise signal. In other words, each different token of noise that may be substituted would give a different internal representation. Therefore, the likelihood that one specific substituted noise token would give an internal representation that is very similar to the internal representation due to the original (unmodified) signal would be very small.

Figure 4 on the other hand shows the main stages of the modified psycho-acoustic model on which the analyser 52 of the preferred embodiment is based. Initially, it will be seen that, for simplicity, the adaptation loops 66 and noise adder 68 of Figure 3 are not employed. However, one or both of these stages can be employed if desired.

However, as distinct from the time-based solution of Dau, the embodiment of Figure 4, transforms the time domain signals produced by the haircell model with transform unit (FFT) 71 into respective frequency domain representations. Then modulation filters 67'are applied in the spectral domain (as a weighting function) to produce a plurality of modulation spectra for each of the x original signals.

In more detail, for each of the x time signals supplied to the transform unit 71 a power spectrum, $R_{fnr}(f)$, for an interval corresponding to about 100 ms of the input signal is calculated. Typically, the noise substituted part (if present) is in the middle of this interval.

For the conversion to modulation spectra (67'), weighting functions $w_{mfnr,fnr}(f)$ are defined where 'mfnr' is the index of the weighting function (or modulation filter number) and 'fnr' is the number of the auditory filter channel from the filterbank 62 and $w_{mfnr,fnr}(f)$ is a function of frequency. For low frequencies the bandwidths of the individuals filters 67' are small and constant (e.g. 10 to 50 Hz) and above a certain frequency the filters have a constant Q preferably between 1 and 4. The shape of the window function can for example be a Hanning window shape, or the amplitude transfer function of a gamma-tone filter. In a preferred implementation, the smallest filter width is 50 Hz, and Q=2. It will be seen that the lowest frequency weighting function is centred at 0 Hz, and so covers only the upper half of the filter shape (everything beyond the maximum).

The weighting functions are squared and multiplied with the power spectra to result in a series of numbers $P_{mfnr,fnr}(f)$ that are used as the internal representation that is fed to an averager 70'.

To illustrate this Figures 5 and 6 show the power spectra ($R_{fnr}(f)$) of an harmonic tone-complex and Gaussian noise respectively provided as input to the filterbank 67'. Figures 9(a) and 9(b) illustrate the input (R_{25}) corresponding to Figures 5 and 6 and modulation spectrum output ($P_{25,18}$) of one of the filters (25,18) of the filterbank 67' for an harmonic tone complex with a fundamental frequency of 100 Hz and for a noise input signal respectively. Both input signals are of equal spectral density and total level. However, it is clear that the filter $P_{25,18}(f)$ has an average higher output level for the harmonic tone complex than for the noise signal. Thus, the summed values ($M_{25,18}$) will be different. For the noise signal M is 0.0054, whereas for the harmonic tone complex M is 0.0093, nearly a factor of two difference. So a matrix of values M presents a representation that differs considerably for noise and harmonic tone complex signals and this shows that classification of noise signals using this model is possible.

In the model of Figure 4, the powers $P_{mfnr,fnr}(f)$ for each modulation spectrum are summed (70') to produce a value for each cell in a matrix M. In this way the activity (M(fnr,mfnr)) within each modulation filter averaged over some time (9 frames) is determined. This average is not sensitive to the specific details of a noise signal which obviates the problem of using the Dau model outlined above. The activity for each filter for one signal can then be compared with the corresponding activity (M') for another signal processed in parallel to provide a perceptual measure D of the difference between the signals:

$$D = \sqrt{\sum_{fnr} \sum_{mfnr} (M - M')^2 / M^2}$$

The value D can then compared to a criterion to determine whether noise substitution is allowed. It should be noted that the criterion can be frequency dependent. For example, for low frequencies, the criterion can be lower and proportional to the bandwidth of the auditory filters; and for high frequencies the criterion can be constant.

Also, the selection component 16' or analyser 52, Figure 2, may require that more than a threshold number of contiguous frequency bands for more than continuous number of intervals can be modelled with noise before instructing the multiplexer (MUX) to switch to a noise model, as only when these thresholds are exceeded would the required saving in bit-rate be made by swapping to a noise model.

In experiments, the embodiment described above was tested on a number of short (300 ms) segments of stationary audio. It was found in a listening test that with 50% to 80% of bandwidth replaced, an audio quality could be obtained that was comparable to that of MPEG 1 Layer III at a bitrate of 96 kbit/sec for mono audio.

In the first embodiment of the invention, noise is iteratively substituted and tested. For each test, the model output of the original signal is compared to the model output of a modified signal i.e. with noise substituted. Based on this comparison a decision is made whether noise can be substituted or not. However, it will be seen that this approach is computationally intensive.

An alternative approach is to make a direct decision for particular time intervals and for particular auditory filters (62,67') that are suspected to be good candidate spectro-temporal intervals for noise substitution, for example, intervals having low energy levels.

In this case one input signal, say I/P2, comprises a synthetic noise signal. The model output (Rep 2) for this signal is then compared directly to the model output (Rep 1) for the original signal to provide a difference measure (D). It will be seen that for a given spectro-temporal interval Rep 2 can be pre-calculated so reducing the computational intensity of this approach.

When the difference between Rep 1 and Rep 2 is smaller than a certain criterion one can assume that noise can be substituted within that particular spectro-temporal interval because apparently in that interval the input audio signal is very similar to a noise signal (in a perceptual sense).

It will be seen that in the first embodiment, masking is inherently taken into account in the decision process. This is useful because when a certain spectro-temporal interval is masked, it can be substituted with noise without any problem. In the alternative implementation, it cannot be seen directly how modification of a certain spectro-temporal interval will affect the model output. In order to be able to do this, it is beneficial to consider to what extent the candidate spectro-temporal interval for noise substitution is masked by other signal components. This can be taken into account by giving a rating to the detectability (det) of the substitution of a spectro-temporal interval, i.e. the degree to which it is masked by other components. So, for example, a low energy interval within a high power signal would have a low detectability rating. The product of detectability (det) and the difference measure (D) that is obtained for an candidate interval is assumed to be a good indicator as to whether noise can be substituted or not.

This approach is much faster than the approach of the first embodiment because it requires only a single pass (instead of many) of the original input signal through the model plus the derivation of the masking properties, something which can be achieved without extensive computational complexity.

It will be seen that the invention is not alone applicable to an MPEG encoder, rather it is applicable in any encoder where a signal is encoded parametrically with noise and by some other means. Referring now to Figure 7, in a second embodiment of the present invention the improved selection component 16" is employed within a parametric audio coder 80 to provide enhanced discrimination between noisy and non-noisy spectro-temporal intervals. An example of such a parametric coder is the sinusoidal description of audio signals, which is highly suitable for various tonal signals, described in European Patent Application No. 02077727.2 filed 8 July 2002 (Attorney No. PHNL020598). Within the coder, a sinusoidal analyser 82 transforms sequential segments of an input signal x(t) into the frequency domain, with each segment or frame then being modelled using a number of sinusoids represented by amplitude, frequency and possibly phase parameters C_S. When the synthesised sinusoidal components of a signal have been removed from the input signal, the residual signal can then be assumed to comprise noise and this is modelled in a noise analyser 84 to produce noise codes C_N. Each of the sinusoidal codes and noise codes C_S, C_N are then encoded in a bitstream AS. Other components of the signal which may be coded include transients and harmonic complexes, however, these are not described here for clarity.

The invention is implemented in such an encoder as follows: The original input signal x(t) is first coded by default to provide a combination of noise and sinusoidal

codes $C_{S(1)}$, $C_{N(1)}$ and these coded segments are provided as input I/P1(0) of a selection component 16" corresponding to the component 16' of Figure 2.

Then for each of a plurality of frequency bands *i* in a given segment *n*, the sinusoidal analyser 82 does not encode sinusoidal components within the frequency band and so the (greater) residual signal is encoded by the noise analyser 84. Each of the candidate noise and sinusoidal codes $C_{S(i)}$, $C_{N(i)}$ produced are then provided to I/P2(i) of the selection component 16". Based on the resulting distortion D, a decision can be made about which candidate set of codes $C_{S(i)}$, $C_{N(i)}$ is most efficient in terms of bitrate and does not have a distortion that exceeds the predefined threshold.

Referring now to Figure 8, as in the first embodiment, for each input I/P1 and I/P2(i), codes for a plurality of segments s1,s2 and s'1(i),s'2(i), are synthesized and combined using respective Hanning window functions in units 42' to provide time-windowed signals for an interval t(n) as inputs to the perceptual analyser 52, which operates as described in relation to the first embodiment. The analyser 52 therefore provides a decision as to whether the modelling of a given band in a given segment with a combination of sinusoids and noise (I/P1) as compared to noise alone (I/P2(*i*)) will be audible or not. It can then be left to the multiplexer 15' to determine which sets of codes 1...*i* to employ across segments ...s1,s2... to provide an optimum bit rate for encoding the signal x(t).

As in the first embodiment, rather than iteratively testing each interval against a noise substituted version of the input signal, a candidate spectro-temporal interval of the input signal can simply be compared against a pre-calculated representation for a noise signal for the same interval to determine whether the candidate interval is noisy or not.

In either case, this means that for a parametric coder, noise-classified intervals need not be represented by sinusoids or other components such as harmonic complexes or transients with possible savings in bit rate and possible quality improvement because a noisy interval would not be represented by sinusoids in particular.

It will be seen that using the second embodiment in particular, the specified spectro-temporal intervals of an audio signal replaced by noise will have an energy equal to that of the conventionally modelled audio signal.

As described above in relation to both embodiments, in order to let the noise substitution work well, it was found that it is important to first substitute noise over a longer temporal interval to determine whether substitution is allowed. After that, the actual final substitution is only done for a much smaller interval. Although the invention may be implemented as such, it has been found that, in general, if noise is only classified in the test

interval that will later be used for the final substitution, rather unreliable classifications will result.

However, if employing long temporal test intervals proves problematic, instead of taking such a long interval for classification, a broad spectral interval (with a short duration) could also be used, with the final substitution only being made in a narrower spectral interval.

CLAIMS:

1. A method of classifying a spectro-temporal interval of an input audio signal (x(t)) comprising:
first modelling (62...71) said spectro-temporal interval of said input audio signal according to a perceptual model to provide a first representation (Rep 1);
second modelling (62...71) said spectro-temporal interval using a modified noise substituted input signal according to said perceptual model to provide a second representation (Rep 2); and
classifying (52) said spectro-temporal interval of said audio signal as being noise or not based on a comparison of said first and second representations.

2. A method according to claim 1 wherein said perceptual model comprises:
a first plurality of *x* filters (62), each providing respective band-pass filtered time-domain signals derived from said input audio signal for each of a first plurality of frequency bands;
a rectifier (63) and a low pass filter (64) for processing each of said band-pass filtered signals;
a transformer (71) for providing a frequency spectrum representation ($R_{fnr}(f)$) of said processed and filtered signals; and
a second plurality of *y* filters (67'), each providing respective band-pass filtered frequency-domain signals ($P_{fnr,mfnr}(f)$) derived from each of said transformed signals for each of a second plurality of frequency bands;
wherein each of said first and second representations comprise an $x*y$ matrix (M, M') of filtered frequency-domain information.

3. A method according to claim 2 wherein each of said first and second representations comprise an $x*y$ matrix including an integral of said filtered frequency-domain information.

4. A method according to claim 1 wherein said modified noise substituted input signal comprises a temporal interval (t(n)) of said input audio signal in which a frequency band (*i*) is replaced with a noise modelled signal.

5. A method according to claim 4 comprising the steps of:
iteratively replacing frequency bands (*i*) of said temporal interval (t(n)) of said input audio signal with a noise modelled signal to provide a series of modified input signals each corresponding to a candidate spectro-temporal interval to be classified;
iteratively modelling said series of modified input signals to provide a series of second representations; and
iteratively classifying said candidate spectro-temporal intervals based on a comparison of said first and each of said series of second representations.

6. A method according to claim 1 wherein said spectro-temporal interval of said input audio signal comprises a selected frequency band for a temporal interval of said input audio signal and wherein said modified noise substituted input signal comprises a noise modelled signal for said frequency band.

7. A method according to claim 6 wherein said second modelling step is performed only once.

8. A method according to claim 6 further comprising the step of:
determining the extent (det) to which substitution of a noise in an input signal for said selected frequency band will be masked by the remainder of the input audio signal and wherein said classifying step (52) comprises classifying said spectro-temporal interval of said audio signal as a function of said comparison of said first and second representations and the extent of said masking.

9. A method of coding an audio signal comprising:
classifying (16',16") a spectro-temporal signal of said audio signal as noise or not according to the steps of claim 1;
modelling (17,84) at least portion of a spectro-temporal interval classified as noise with noise model parameters; and
encoding (15,15') said noise model parameters in a bit stream (AS).

10. A method according to claim 9 wherein said portion of a spectro-temporal interval comprises a temporal sub-set of said spectro-temporal interval.

11. A method according to claim 9 wherein said portion of a spectro-temporal interval comprises a spectral sub-set of said spectro-temporal interval.

12. A method according to claim 9 wherein said spectro-temporal interval comprises a time period of greater length than a basic interval length (s1,s2) in said bit stream.

13. A component for classifying a spectro-temporal interval of an input audio signal (x(t)) comprising:
means for modelling (62...71) said spectro-temporal interval of said input audio signal according to a perceptual model to provide a first representation (Rep 1);
means for modelling (62...71) said spectro-temporal interval using a modified noise substituted input signal according to said perceptual model to provide a second representation (Rep 2); and
means classifying (52) said spectro-temporal interval of said audio signal as being noise or not based on a comparison of said first and second representations

14. A coder including a component according to claim 13 wherein said component is employed to determine if a spectro-temporal interval is to be coded using noise model parameters.

15. A coder according to claim 14 wherein said coder is one of a sinusoidal coder or an MPEG type coder.

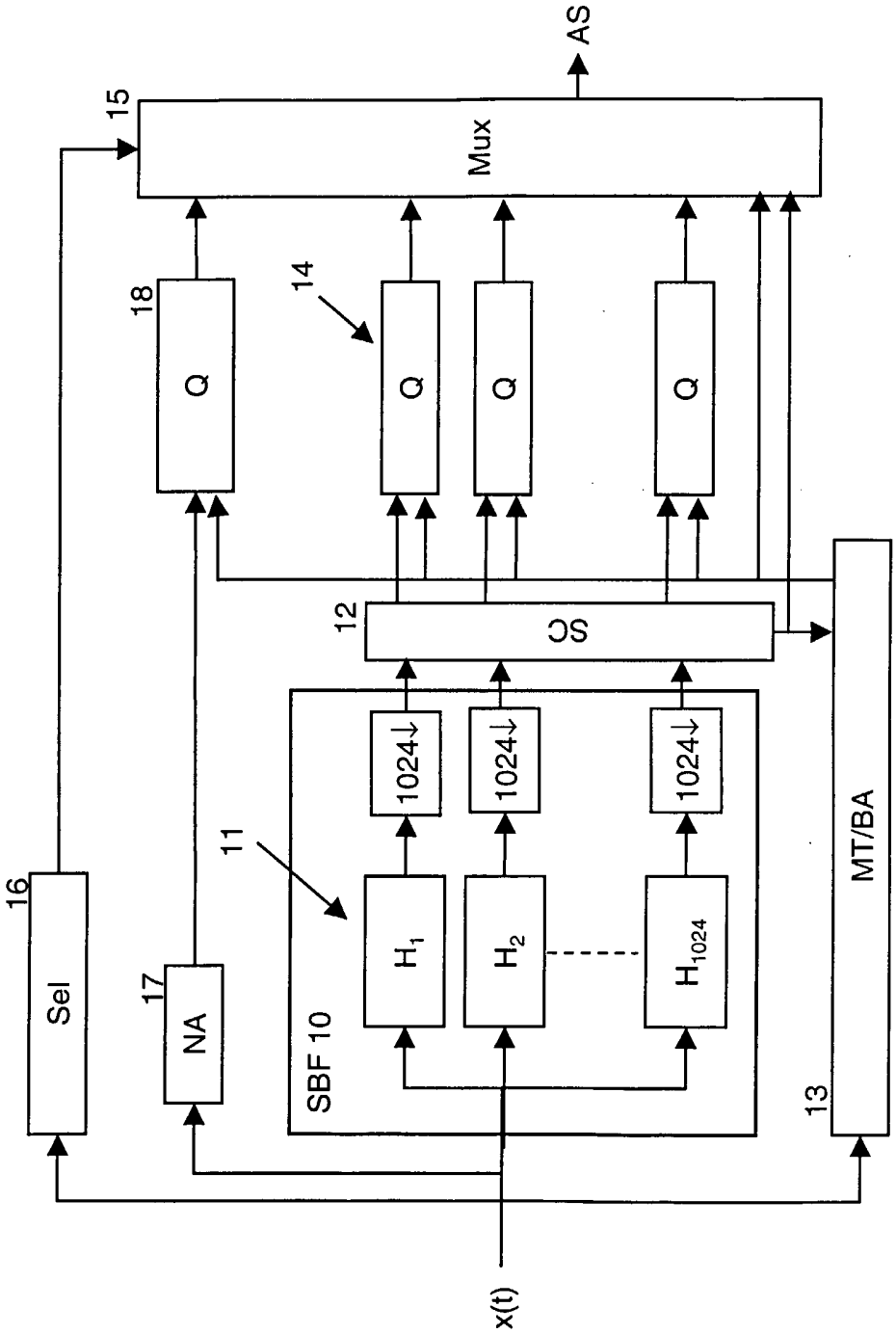
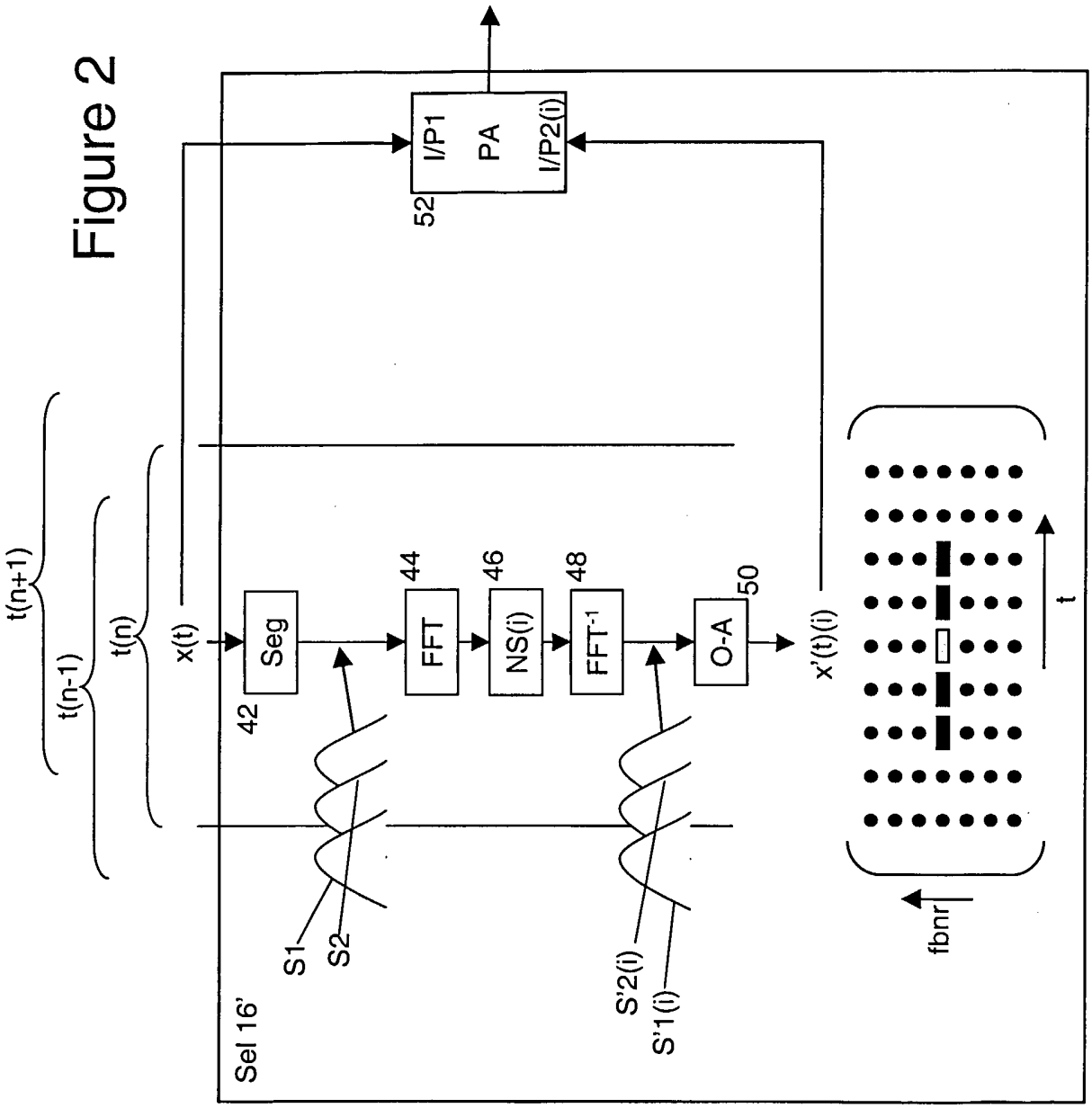

AS
15
Mux
18
14
Q
Q
Q
Q
12
CS
1024↓
1024↓
1024↓
11
H_1
H_2
H_{1024}
SBF 10
MT/BA
13
16
Sel
17
NA
x(t)

Figure 1

Figure 2


I/P1
PA
I/P2(i)
52
t(n+1)
t(n)
t(n-1)
x(t)
Seg
42
FFT
44
NS(i)
46
FFT-1
48
O-A
50
x'(t)(i)
t
fbnr
S1
S2
S'2(i)
S'1(i)
Sel 16'

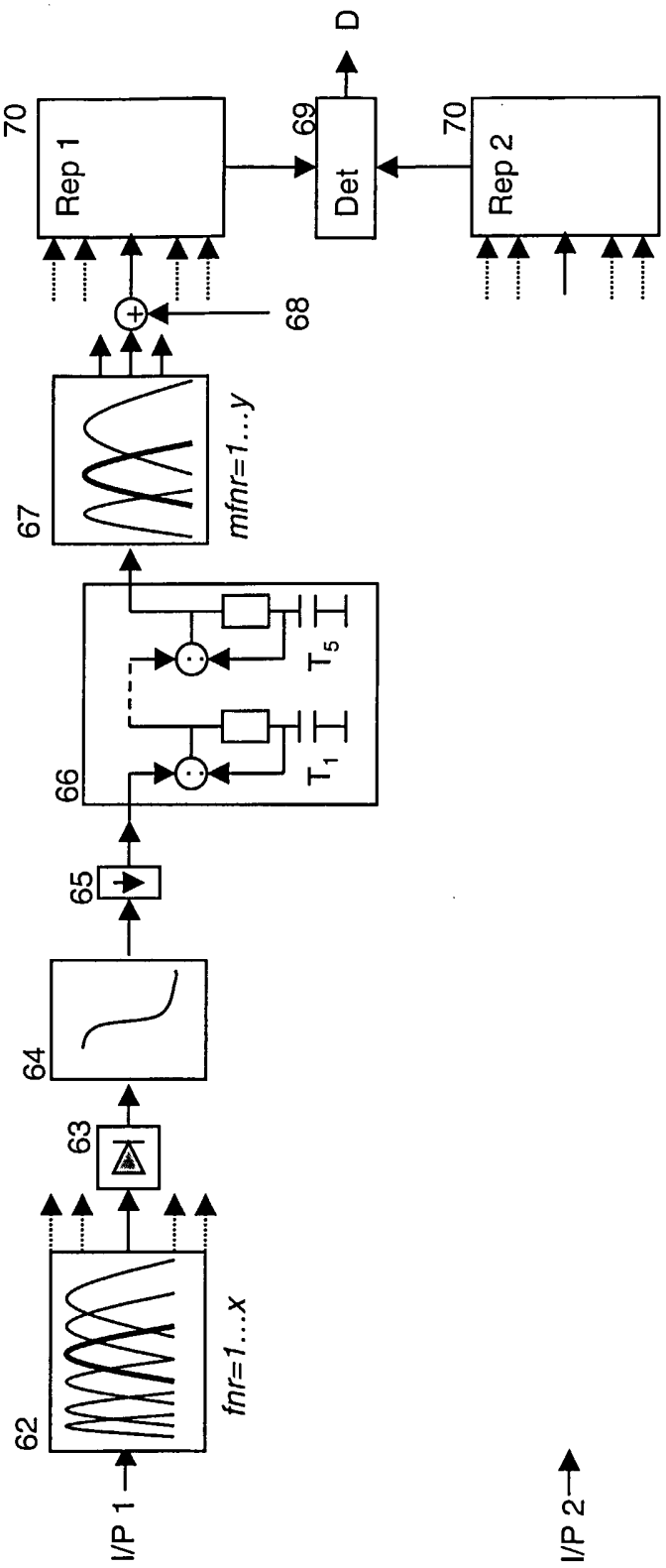

D
70
Rep 1
69
Det
70
Rep 2
68
67
mfnr=1...y
66
T5
T1
65
64
63
62
fnr=1...x
I/P 1
I/P 2

Figure 3

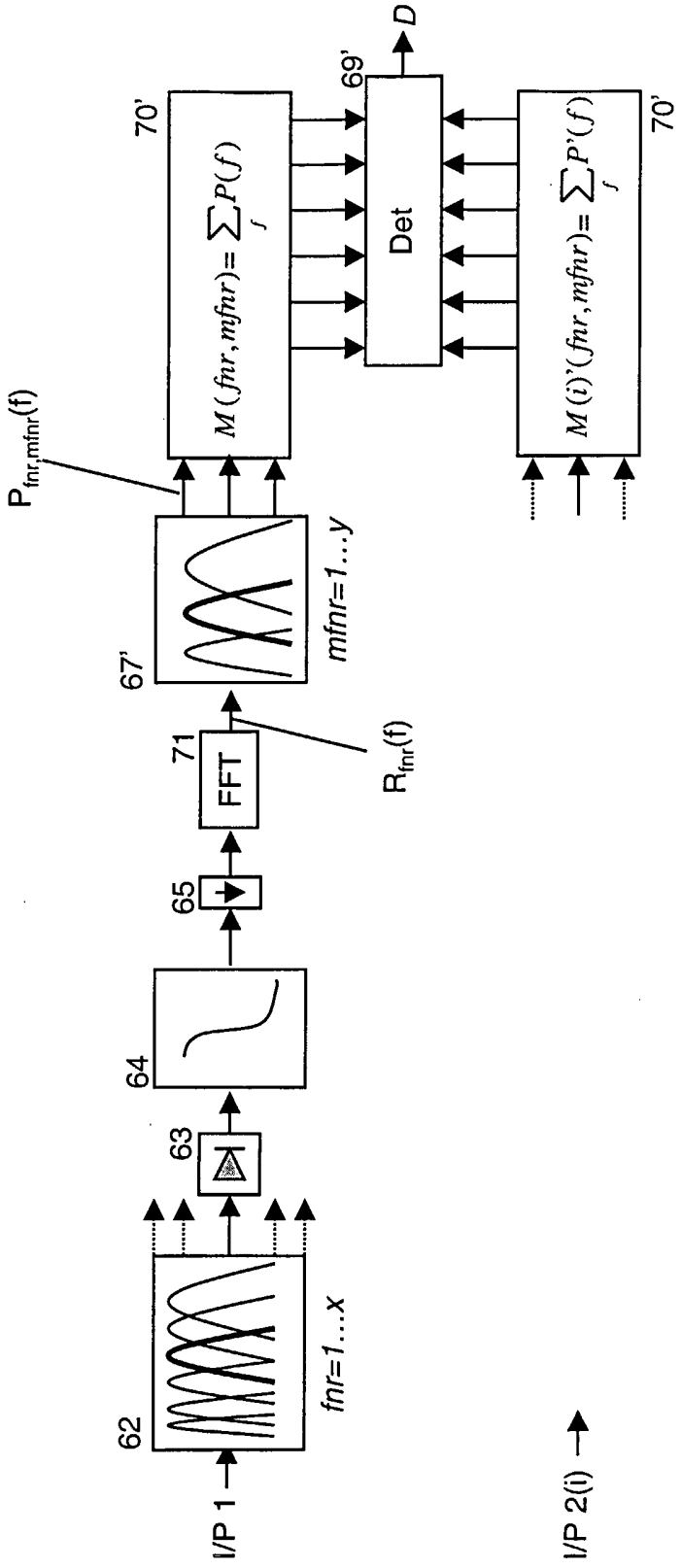

D
69'
70'
Det
$M(fnr, mfnr) = \sum_f P(f)$
$M(i)'(fnr, mfnr) = \sum_f P'(f)$
70'
$P_{fnr,mfnr}(f)$
67'
mfnr=1...y
71
FFT
$R_{fnr}(f)$
65
64
63
62
fnr=1...x
I/P 1
I/P 2(i)

Figure 4

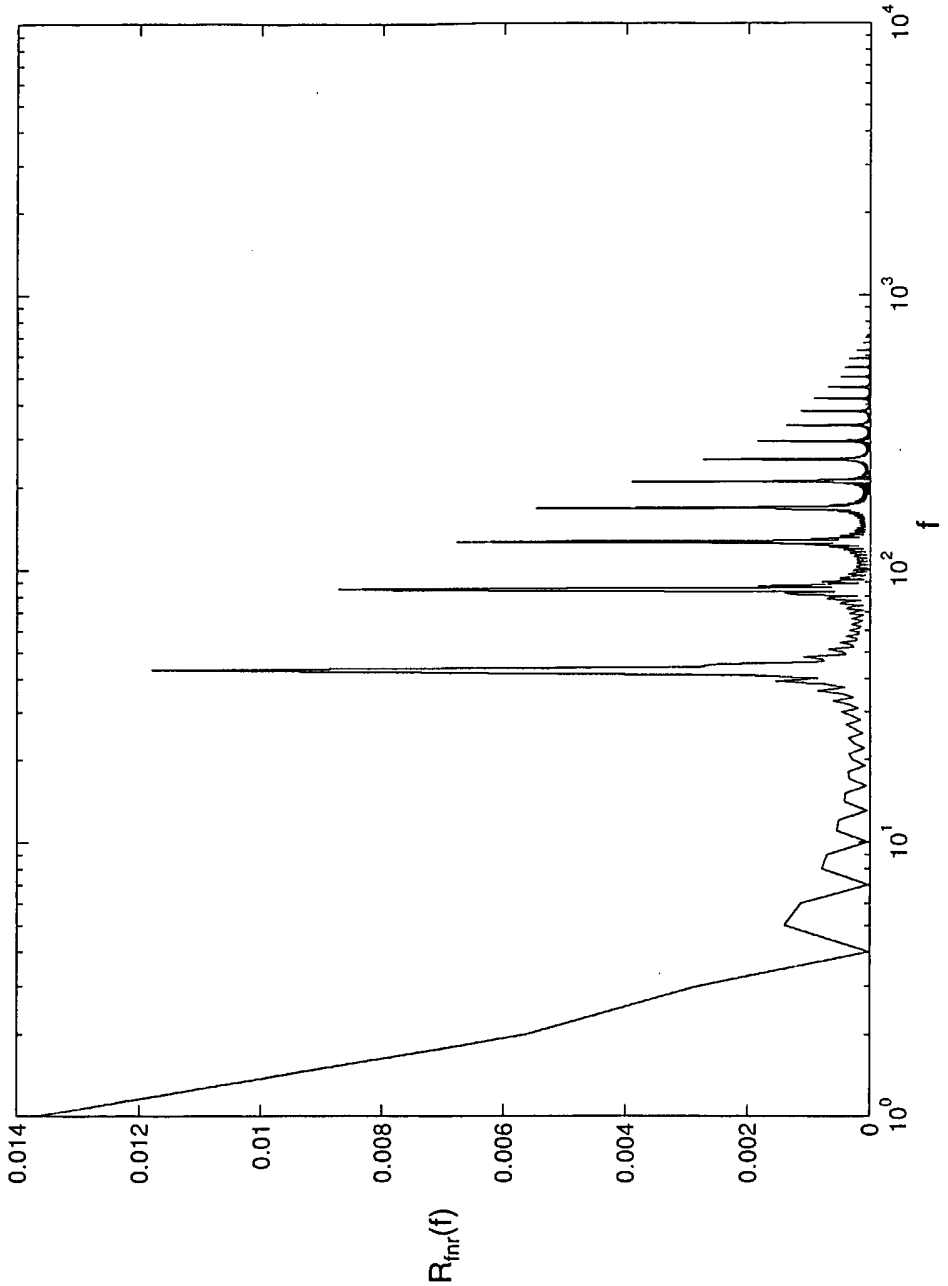


Figure 5

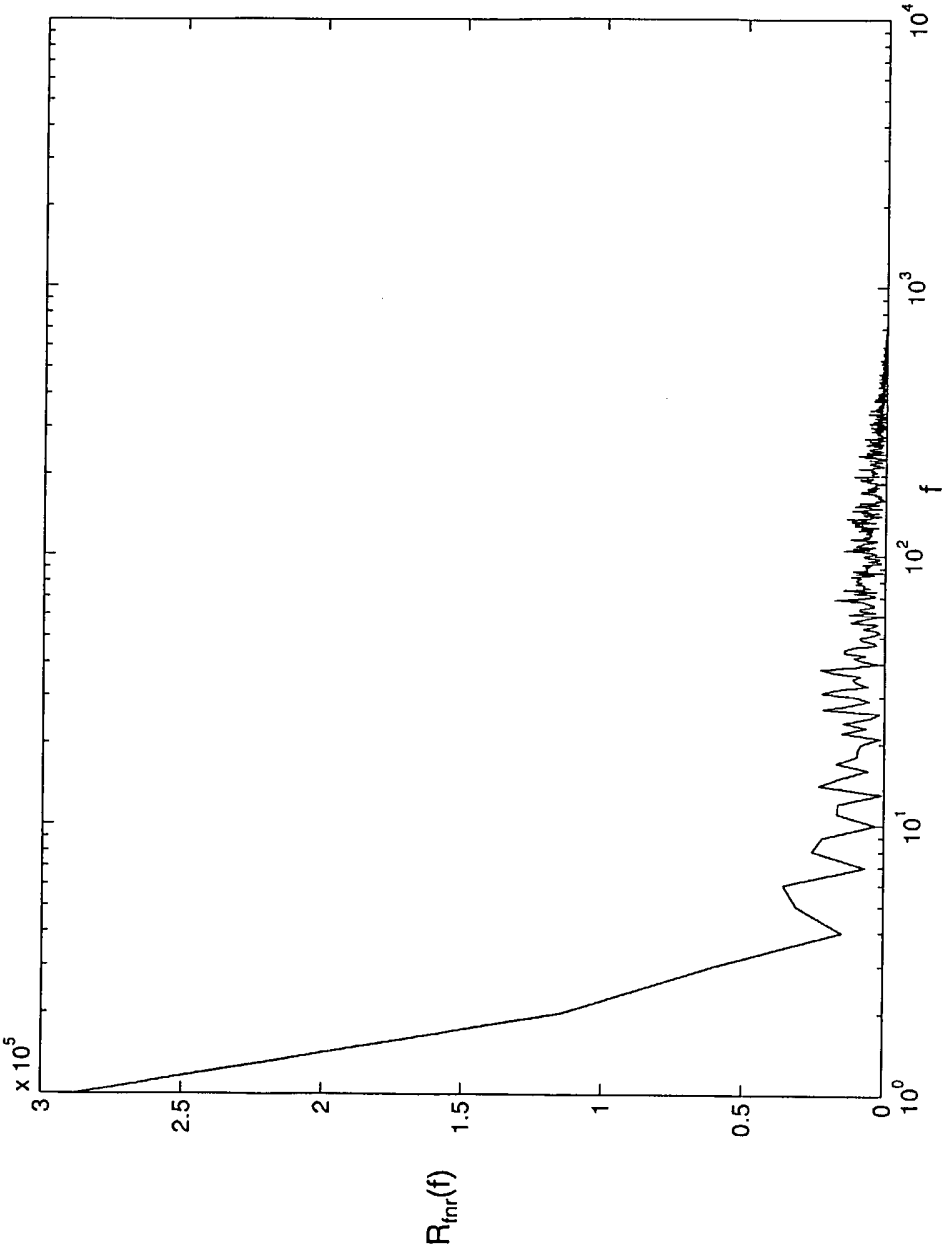


Figure 6

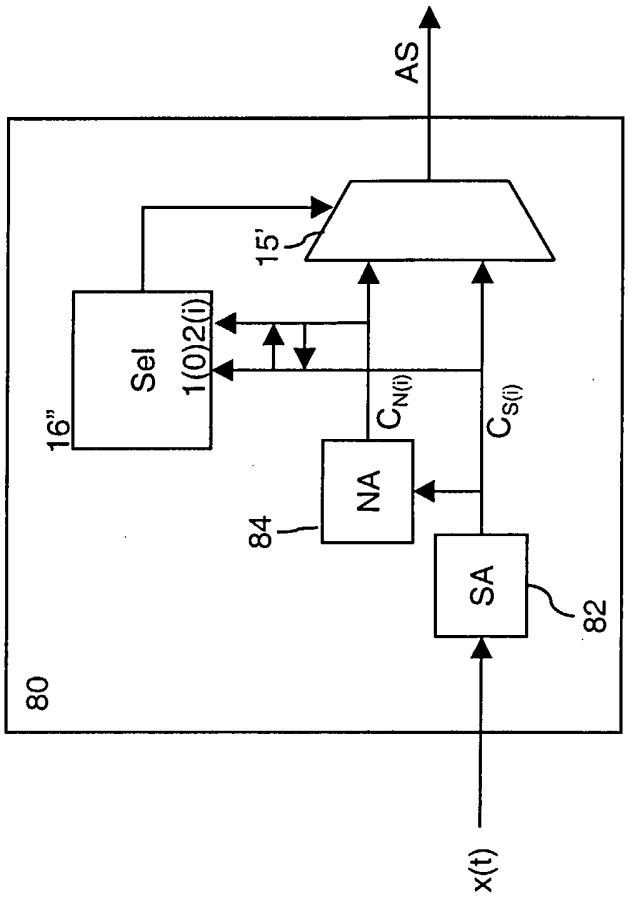
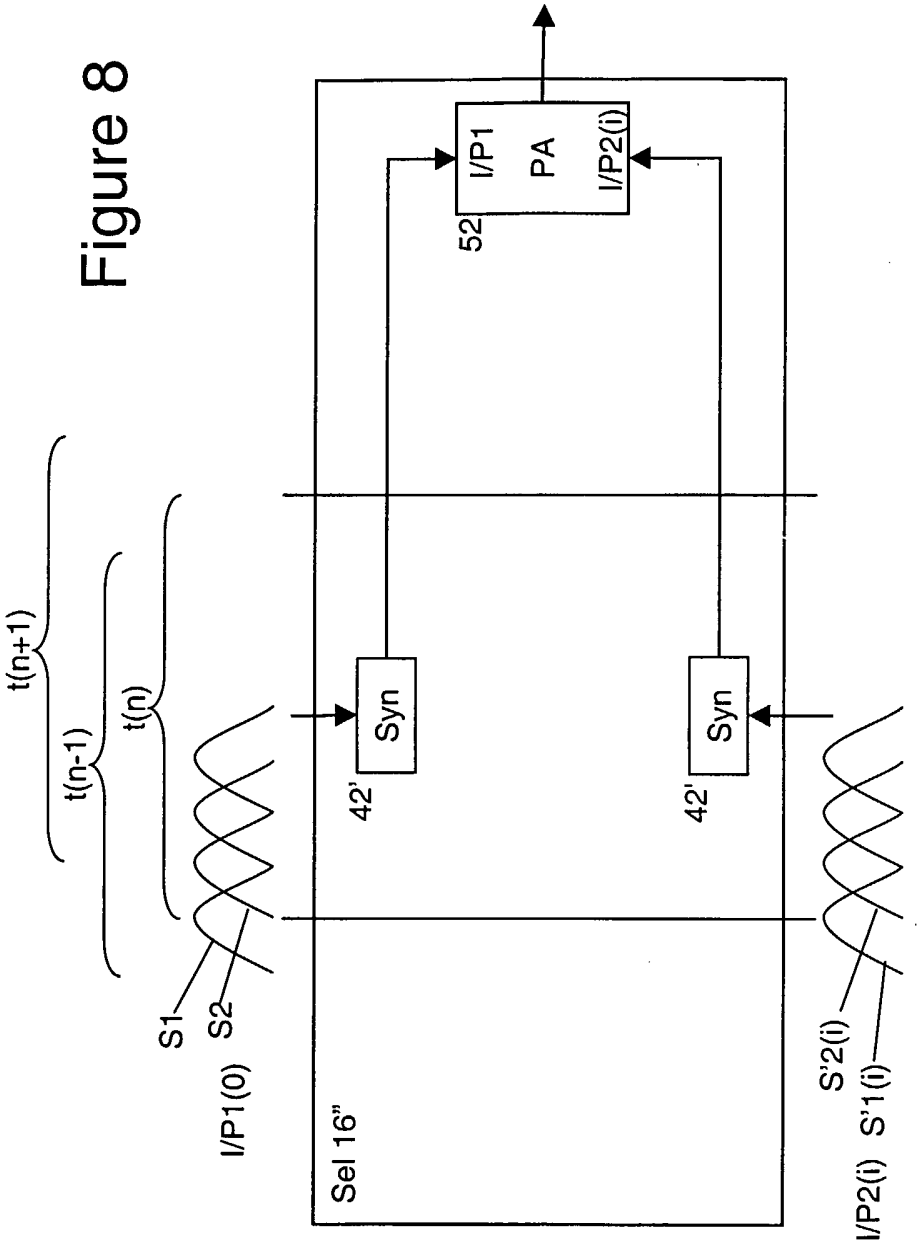
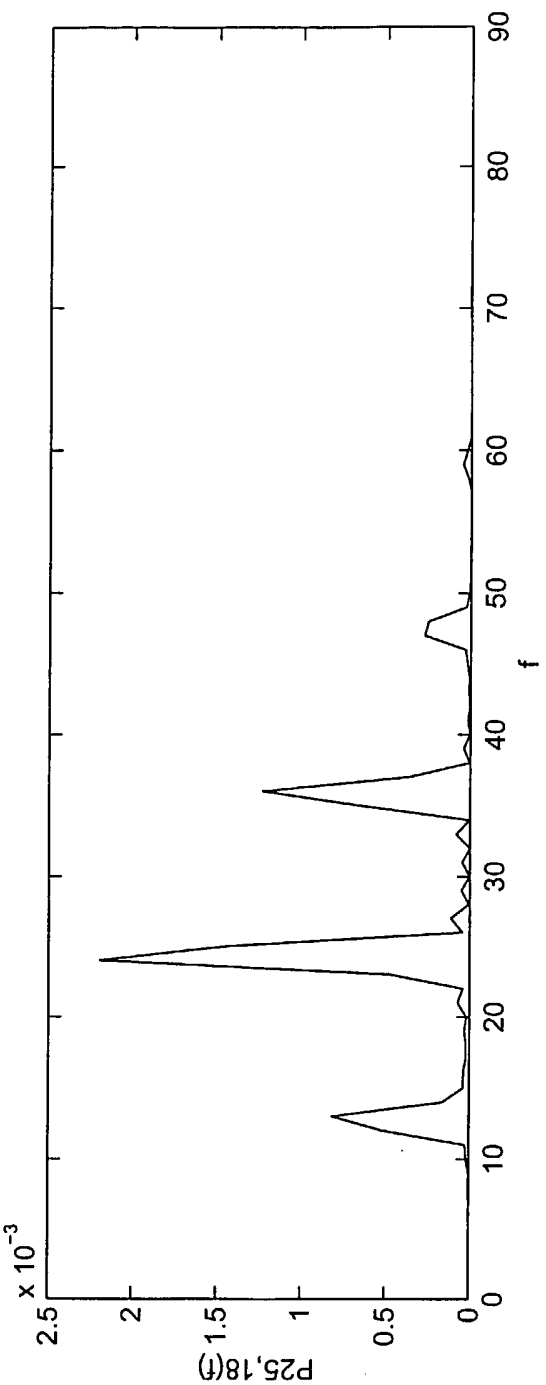


Figure 7

Figure 8


I/P1
PA
I/P2(i)
52
Syn
42'
Syn
42'
t(n+1)
t(n-1)
t(n)
S1
S2
I/P1(0)
Sel 16"
S'2(i)
S'1(i)
I/P2(i)


x 10^-3
2.5
2
1.5
1
0.5
0
P25,18(f)
0
10
20
30
40
50
60
70
80
90
f

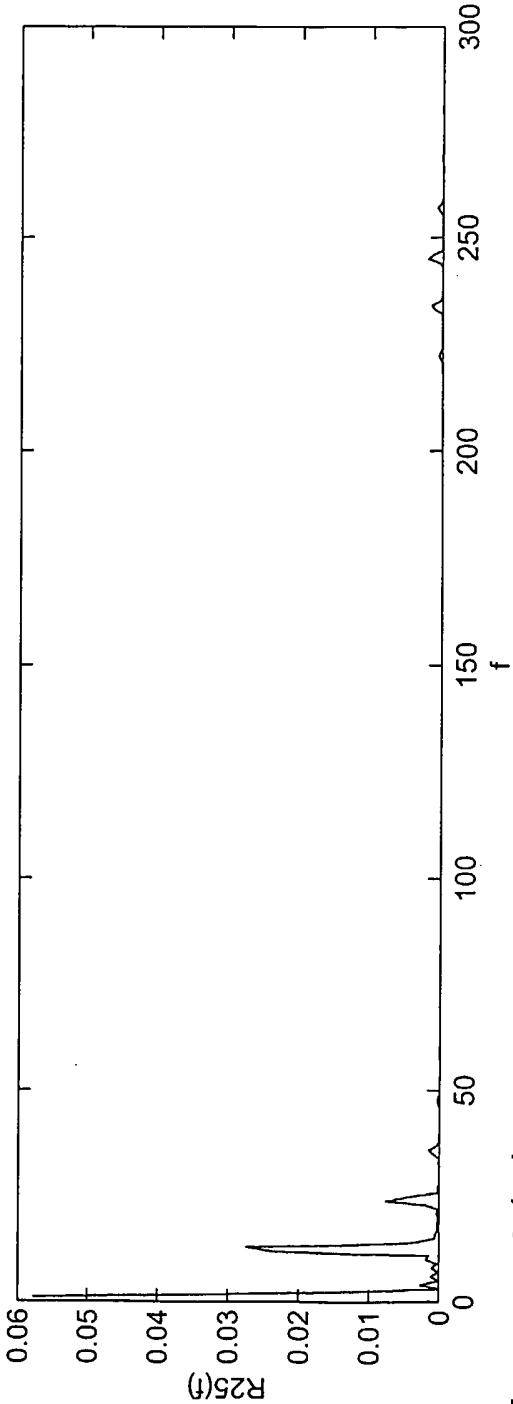

0.06
0.05
0.04
0.03
0.02
0.01
0
R25(f)
0
50
100
150
200
250
300
f

Figure 9(a)

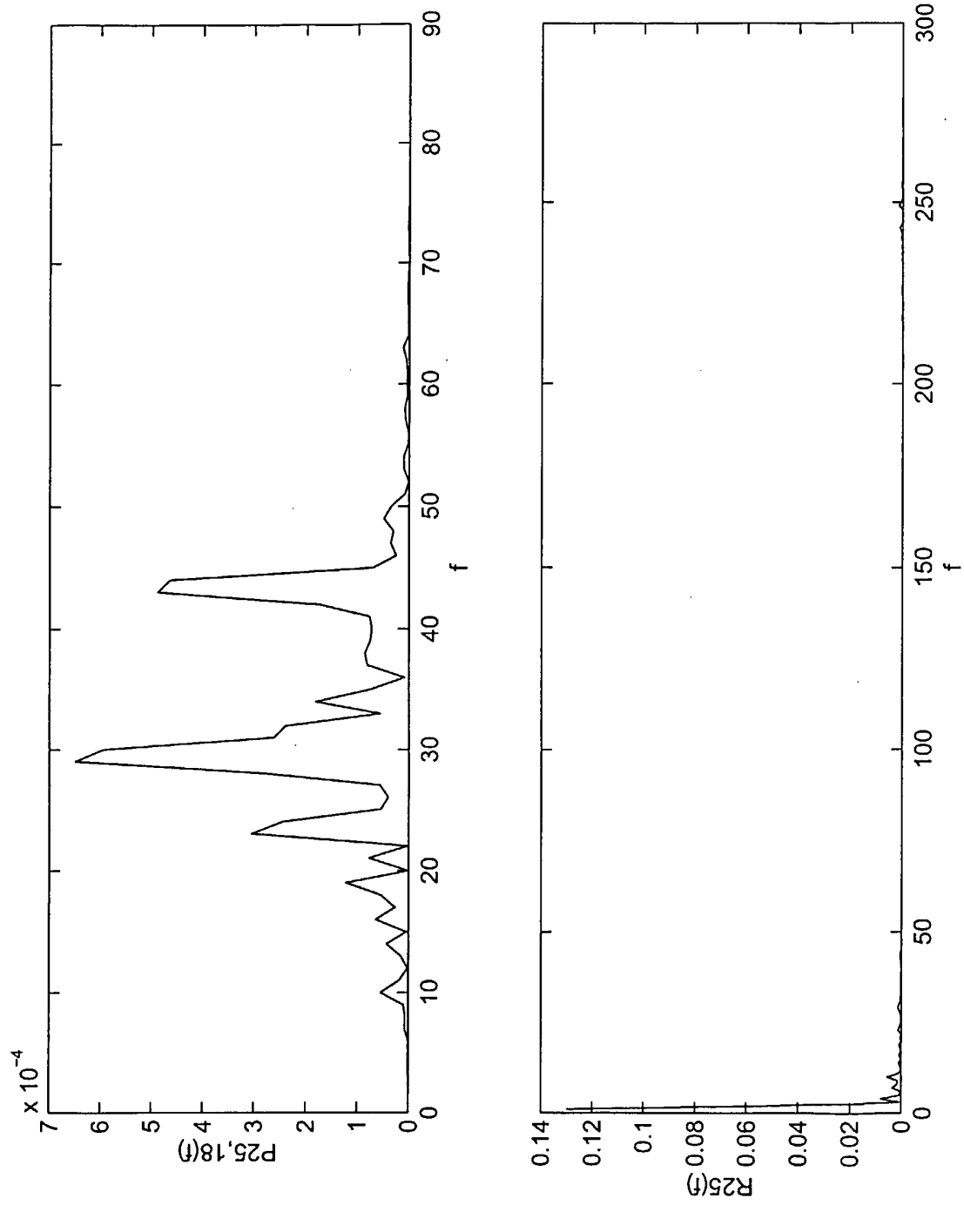


Figure 9(b)

INTERNATIONAL SEARCH REPORT

Interr Application No
PCT/IB 03/02336

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L19/14 G10L11/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, INSPEC, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LEVINE S N ET AL: "Improvements to the switched parametric and transform audio coder" APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 1999 IEEE WORKSHOP ON NEW PALTZ, NY, USA 17-20 OCT. 1999, PISCATAWAY, NJ, USA,IEEE, US, 17 October 1999 (1999-10-17), pages 43-46, XP010365091 ISBN: 0-7803-5612-8 page 44, right-hand column, paragraph 3.2.1 page 44, right-hand column, paragraph 3.2.4 page 45, left-hand column, paragraph 3.2.4 --- -/--	1,9,13, 14

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 October 2003	29/10/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Ramos Sánchez, U

INTERNATIONAL SEARCH REPORT

Intern al Application No.
PCT/IB 03/02336

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	HERRE J ET AL: "EXTENDING THE MPEG-4 AAC CODEC BY PERCEPTUAL NOISE SUBSTITUTION" PREPRINTS OF PAPERS PRESENTED AT THE AES CONVENTION, XX, XX, 1998, pages 1-14, XP008006769 page 2, paragraph 2 -page 3, paragraph 3 ---	1,9,13, 14
A	SCHULZ D: "IMPROVING AUDIO CODECS BY NOISE SUBSTITUTION" JOURNAL OF THE AUDIO ENGINEERING SOCIETY, AUDIO ENGINEERING SOCIETY. NEW YORK, US, vol. 44, no. 7/8, 1 July 1996 (1996-07-01), pages 593-598, XP000733647 ISSN: 0004-7554 cited in the application page 596, right-hand column, paragraph 3 -page 597, left-hand column, paragraph 2 ---	1,9,13, 14
A	VAN DE PAR S ET AL: "A new psychoacoustical masking model for audio coding applications" 2002 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS (CAT. NO.02CH37334), PROCEEDINGS OF INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (CASSP'02), ORLANDO, FL, USA, 13-17 MAY 2002, pages II-1805-8 vol.2, XP002256785 2002, Piscataway, NJ, USA, IEEE, USA ISBN: 0-7803-7402-9 abstract page 1806 -page 1807, left-hand column -----	1,9,13, 14