

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
1 November 2007 (01.11.2007)

PCT

(10) International Publication Number  
**WO 2007/121778 A1**

(51) International Patent Classification:  
G10L 19/02 (2006.01)

(21) International Application Number:  
PCT/EP2006/009601

(22) International Filing Date: 4 October 2006 (04.10.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/745,499 24 April 2006 (24.04.2006) US

(71) Applicant (for all designated States except US): **NERO AG** [DE/DE]; Im Stoeckmaedle 18, 76307 Karlsbad (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **DIMKOVIC, Ivan** [YU/DE]; Rintheimer Str. 8, 76131 Karlsruhe (DE). **PASCUTTO, Gian Carlo** [BE/BE]; Nestor de Tierestraat 31 0202, B-9700 Oudenaarde (BE).

(74) Agents: **ZINKLER, Franz** et al.; P. O. Box 246, 82043 Pullach/Munich (DE).

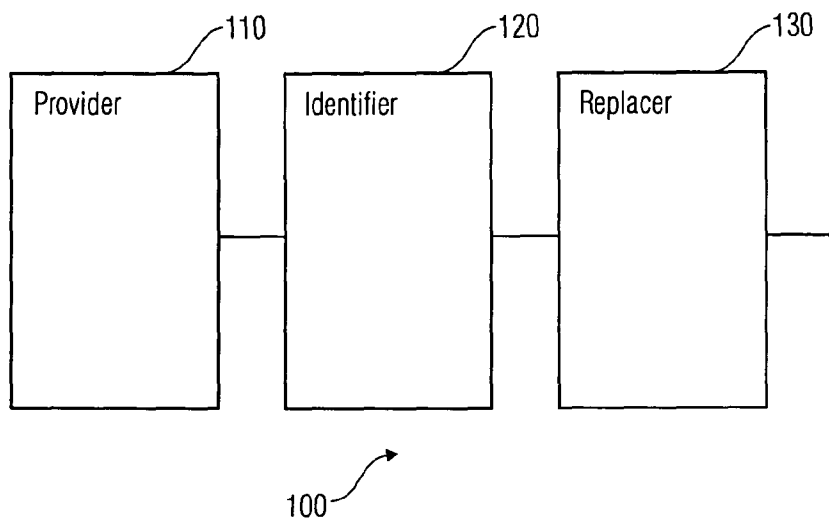
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (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: ADVANCED AUDIO CODING APPARATUS



(57) Abstract: A method and an apparatus for encoding digital audio data with reduced bit rates, the apparatus comprising a provider of psycho-acoustically quantized digital audio data with a bit rate being higher than the reduced bit rate. The apparatus further comprises an identifier for identifying a frequency band according to a selection criterion, the selection criterion being such that an impact on the quality of the digital audio data when the data in the identified frequency band is replaced by generated noise is smaller than the impact on the quality of the digital audio data, which would arise when the data in a different frequency band is replaced by generated noise. The apparatus further comprises a replacer for replacing data in the identified frequency band of the digital audio data by a noise synthesis parameter, the noise synthesis parameter requiring a smaller amount of data than the data in the identified frequency band, the digital audio data having the reduced bit rate.

WO 2007/121778 A1



Unfortunately, at very low bit rates, i.e. at high compression demands, coding systems have no options, but to shut down frequency bands, i.e. replace them with silence.

5 This method is utilized in order to meet the coding demands imposed to the codec. This introduces holes in the spectrum that are especially annoying and they are the biggest contributor to audio coding artifacts.

10 Fig. 8 shows a typical state of the art audio encoder for an input signal that is PCM (Pulse Code Modulation) encoded and input to a filter bank 810 and a perceptual model 815. The input signal is transformed from the temporal or time domain to the frequency domain by the filter bank 810,  
15 which is usually based on well known signal transform functions, such as the modified discrete cosine transform. The outputs of the filter bank are frequency coefficients.

At the same time the signal is evaluated by the perceptual  
20 model 815, the perceptual model evaluates the input signal by mathematically modeling the human auditory system and outputs a measure, such as for example the just noticeable distortion (JND) in units of a signal-to-mask ratio (SMR) of the input signal energy to the just noticeable  
25 distortion or noise energy.

The perceptual model block 815 and the remaining blocks in the state of the art encoder, as it is depicted in Fig. 8, treat the output of the filter bank block 810  
30 proportionally to the critical bandwidths of the human auditory system, for example, by a grouping of the frequency coefficients in so-called scaling factor bands. A good summary of the perceptual model can be found in T. Painter and A. Spanias, "Perceptual Coding of Digital  
35 Audio", in the proceedings of the IEEE, pp. 451-513, April 2000.

The target compression demand is met by quantization of the frequency coefficients. Before quantization, the coefficients are scaled by so-called scaling factors to determine the eventual precision of the quantization process. The bit/noise allocation block 820 is responsible for estimation or calculation of the scaling factors, so the reconstruction of the quantized values yields quantization noise just below the masking threshold estimated by the perceptual model. Under certain circumstances, the perceptual model 815 indicates that certain frequency bands are noise-like and may be modeled by generating noise with a certain energy on the decoder side. For these frequency bands, there is no need to determine scaling factors or frequency coefficients, but parameters for a noise generator at the decoder side are inserted instead. Since the parameters for the noise generator take up less amount of data than scaling factors and frequency coefficients, data rates can be saved by replacing frequency bands with generated noise. The impact of the replacement on the quality of the decoded audio signal is kept in boundaries, determined by the perceptual model. For example, a frequency band, which is to be replaced, must not exceed a certain tonality threshold, nor does it contain any transient signal. The thresholds that determine noise substitution depend on the perceptual model. In ISO/IEC 14496, for example, perceptual noise substitution as a feature of AAC is described.

An advanced coding method used in some perceptual codecs is the so-called perceptual noise substitution (PNS) of which a good summary can be found in Herrer, Jürgen, Schultes, Donald, "Extending the MPEG-4AAC Codec by Perceptual Noise Substitution", AES document 4720.

After the bit allocation block 820 in Fig. 8, quantization is done in the quantization block 825, yielding quantized frequency coefficients, which are brought to the irrelevancy reduction block 830. The irrelevancy reduction

block 830 employs signal irrelevance reduction methods, which are well known from signal theory. For example, Huffman coding, vector quantization or arithmetic coding are well known methods for signal irrelevancy reduction. An  
5 overview of these methods can, for instance, be found in K. Brandenburg, "MP3 and AAC Explained" in proceedings of the AES 17<sup>th</sup> International Conference on High-Quality Audio Coding, 1999.

10 In order to achieve the target coding requirements, for example, a given bit rate for the compressed signal, state of the art codecs are able to reduce the coding requirements by increasing the allowed amount of noise specified by the psycho-acoustic model or perceptual model.  
15 Referring to Fig. 8, the coding requirement is verified in block 835 and if the coding requirement is not met, the bit demand is further reduced in the reduction block 840, upon which the encoding algorithm returns to the bit/noise allocation block 820. If the coding requirement is  
20 achieved, a bit stream multiplexer block 845 multiplexes the coded quantized frequency coefficients and the coded scaling factors into a coded bit stream.

If the coding requirement is not met and the bit demand is  
25 further reduced, additional noise is introduced to the signal. As allowed noise is increased, the scaling factors are increased as well and resolution of the quantized signal is decreased, which then also decreases the bit demand. The quantization resolution can be decreased up to  
30 the point when noise gets greater than the signal itself, possibly meaning the output of the quantization block for that scaling factor will be zero. This effectively inserts a hole in the spectrum in the place where the signal of the scaling factor should be present. This operation can be  
35 iteratively repeated as long as the transmission/storing demand of the coded quantized coefficient is below the constraints imposed to the encoder. This operation always

terminates successfully, even if it sets all quantized outputs to zero, cf. the flowchart in Fig. 8.

While, with the above-described state of the art method the coding requirements are effectively maintained and it functions quite well, provided that the constraints opposed to the codec are achievable without eliminating too much of scaling factors in the constraint's reduction phase, the method could fail miserably if the coding demands are set to be too high for the encoder.

This usually happens if the bit rate required is well below the requirements of the perceptual model. Non-optimized codecs would usually introduce high amounts of holes due to the shut-down of too much scaling factors in order to meet the coding constraints. Spectral holes or shut-downs are usually easily detectable by listeners and they have a huge impact on degradation of the sound quality. Signals containing spectral holes are usually classified as ringing, a swishy sound, birdies, etc.

Optimized state of the art codecs, as they can, for example, be found in 3GPP (3GPP = Third Generation Partnership Project), TS (TS = Technical Specification) 26.403, employ more advantageous strategies of coding constraints reduction, usually called hole avoidance. This strategy works by imposing maximum constraint reduction limits for each scaling factor. This ensures that no holes would be introduced in the scaling factors as long as it would be possible to reduce coding constraints for all scaling factors without violating this limit and maintaining the constraints imposed to the encoder. However, even with this advanced strategy, it is quite possible that the coding constraints will not be met and, in this case, the encoder will have no other option, but to start introducing spectral holes by eliminating scaling factors.

Fig. 9 shows spectrum plots of two codec signals, in the range of 100 Hz to 15 kHz. The codecs displayed are 32 kbps, which corresponds to a 44:1 compression ratio and 320 kbps, which corresponds to a 4.4:1 compression ratio. As it can easily be seen from Fig. 9, the 32 kbps codec was forced to introduce spectral holes in order to meet a coding demand and it can be seen by severe degradations in the upper frequency range.

10

### Summary of the Invention

It is the object of the present invention to provide an apparatus and a method for encoding digital audio data with a reduced bit rate, without introducing spectral holes into the signal.

This object is achieved by an apparatus for encoding digital audio data with a reduced bit rate, comprising a provider of psycho-acoustically quantized digital audio data with a bit rate being higher than the reduced bit rate and an identifier for identifying a frequency band according to a selection criterion, the selection criterion being such that an impact on the quality of the digital audio data when the data in the identified frequency band is replaced by the generated noise is smaller than the impact on the quality of the digital audio data, which would arise when data in a different frequency band is replaced by generated noise. The apparatus further comprising a replacer for replacing data in the identified frequency band of the digital audio data by a noise synthesis parameter, the noise synthesis parameter requiring a smaller amount of data than the data in the identified frequency band, the digital audio data having the reduced bit rate.

35

This object is further achieved by a method for encoding digital audio data with a reduced bit rate, comprising the step of providing psycho-acoustically quantized digital

audio data with a bit rate being higher than the reduced  
bit rate and identifying a frequency band according to a  
selection criterion and the selection criterion being such  
that an impact on a quality of the digital audio data when  
5 the data in the identified frequency band is replaced by  
the generated noise is smaller than the impact on the  
quality of the digital audio data, which would arise when a  
data in a different frequency band is replaced by generated  
noise. The method further comprising a step of replacing  
10 data in the identified frequency band of the digital audio  
data by a noise synthesis parameter, the noise synthesis  
parameter requiring a smaller amount of data than the data  
in the identified frequency band, the digital audio data  
having the reduced bit rate.

15

The present invention is based on the finding that since  
the human auditory system is not able to distinguish  
between different kinds of narrow band signals and noise  
signals as long as the average energy is the same or  
20 comparable. Under some circumstances, where high data  
compression is needed, the quality of digital audio data  
can be preserved more effectively if noise generators are  
used instead of shutting down frequency bands completely.  
This effectively means that it is sufficient to generate  
25 noise at the decoder stage without the need for  
transmitting a quantized spectral coefficient of the scale  
factor band, which is found to be noise-like. The only  
information that needs to be transmitted is the average  
energy value or a noise generator parameter as, for  
30 example, a noise synthesis parameter, of the scale factor  
band, which some codecs, such as MPEG-4AAC transmits  
instead of scaling factor values for such bands if the  
perceptual model indicates its suitability. However, if  
higher compression rates are required, these codecs shut  
35 down frequency bands where further introduction of  
generated noise yields a better quality of the digital  
audio data.



5

Brief Description of the Figures

Embodiments of the present invention will be described using the Figs. attached in which:

- 10 Fig. 1 shows a block diagram of an embodiment of an apparatus for encoding digital audio data;
- Fig. 2 shows a block diagram of a further embodiment of an apparatus for encoding digital audio data;
- 15 Fig. 3 shows an embodiment of an inventive provider;
- Fig. 4 shows a block diagram of another embodiment of an apparatus for encoding digital audio data;
- 20 Fig. 5 shows a flowchart of an embodiment of a sequence controller method;
- Fig. 6 shows a flowchart of an embodiment of an analysis-by-synthesis method;
- 25 Fig. 7 shows a flowchart of an embodiment of a state of the art method extended by an embodiment of the inventive method;
- 30 Fig. 8 shows a flowchart of the state of the art encoding process; and
- 35 Fig. 9 shows two spectral diagrams of encoded digital audio data.

Detailed Description of the Invention

An embodiment of an apparatus 100 for encoding digital audio data with reduced bit rate is depicted in Fig. 1. The embodiment depicted in Fig. 1 comprises a provider 110, which provides psycho-acoustically quantized digital audio data of a bit rate being higher than the reduced bit rate to an identifier 120. The identifier 120 identifies a frequency band according to a selection criterion, the selection criterion being such that an impact on the quality of the digital audio data when the data in the identified frequency band is replaced by generated noise is smaller than the impact on the quality of the digital audio data, which would arise when data in a different frequency band is replaced by generated noise. The identifier 120 indicates the identified frequency band to a replacer 130. The replacer 130 replaces data in the identified frequency band of the digital audio data by a noise synthesis parameter, the noise synthesis parameter requiring a smaller amount of data than the data in the identified frequency band, so that the digital audio data has a reduced bit rate.

A further embodiment of the apparatus 100 for digital audio data is depicted in Fig. 2. Fig. 2 shows the provider 110, the identifier 120 as well as the replacer 130, as they were described with respect to Fig. 1. Furthermore, the embodiment of the apparatus 100 for encoding digital audio data depicted in Fig. 2 comprises an entropy encoder 140 for encoding digital data with a reduced bit rate. The two embodiments depicted in Figs. 1 and 2 of the apparatus 100 can be operative to encode digital raw data, for example, PCM data (PCM = Pulse Code Modulation). The provider 110 can, therefore, be implemented as any source of audio data, for example, a CD player, extended by a means for realizing the psycho-acoustically encoding. The psycho-acoustically encoding is done per frequency band, which can be implemented, for example, by employing a filter in a filter bank within the provider. According to the embodiment

depicted in Fig. 2, the apparatus 100 can comprise an entropy encoder 140, so the digital audio data with reduced bit rates can be entropy encoded, for example, with a Huffman code in order to comply with AAC or MP3 standards.

5

Fig. 3 shows an embodiment of the provider 110. In this embodiment, the provider 110 comprises a filter bank 112, which transforms digital audio data into the frequency domain providing frequency coefficients per frequency band.

10 The provider 110 further comprises a scale factor quantization and noise substitution block 114, which determines the scale factors and the quantization as well as the noise substitution based on the data, a psycho-acoustic model and pre-analyzer block 116 derived from the

15 input digital audio data. The psycho-acoustic model and pre-analyzer block 116 determines from the digital input data as to which frequency bands can be substituted by noise right away and provides that information to the scale factor quantization and noise substitution block 114.

20 Furthermore, the psycho-acoustic model provides data that allows for derivation of the scaling factors and the quantization. The pre-analyzer could analyze the data in the time domain and in another embodiment, it could analyze the data in the frequency domain in order to determine

25 frequency bands that can be replaced with noise at a decoder. One method in order to determine these frequency bands is an analysis-by-synthesis, where basically all frequency bands are sequentially substituted by noise, the complete signal is synthesized again and a quality measure

30 is taken. Running an iteration across all of the frequency bands can identify a frequency band with the lowest quality impact, which would then be chosen for replacement. This process will be detailed later on.

35 In another embodiment of the present invention, the provider 110 would acquire already-encoded data, for example, an MP3 file or AAC encoded data and would then utilize a decoder in order to remove the entropy coding.

Once the entropy coding is removed, psycho-acoustically quantized data that may already contain noise replaced frequency bands, is available to be passed on by the provider 110 to the identifier 120. It is then a task of the identifier 120 to identify the frequency bands, pass on the psycho-acoustically quantized data to the replacer 130, where the according frequency bands are replaced.

In another embodiment, the apparatus 100 is required to reduce the bit rate of digital audio data to a certain target bit rate. An embodiment for this inventive apparatus 100 is depicted in Fig. 4. Fig. 4 shows again an embodiment of the apparatus 100 for encoding digital audio data, which is, at first, provided by a provider 110. The identifier 120 identifies the frequency bands, which are to be replaced by the replacer 130, where the identification is based on a selection criterion. The apparatus 100 in Fig. 4 further comprises a sequence controller 150, which is coupled to the identifier 120 and the replacer 130. Once a frequency band has been identified, the replacer 130 replaces the data in this frequency band by a synthesis parameter for the noise generator, upon which a new bit rate results. It is now the objective of the sequence controller 150 to adjust the selection criterion for the frequency bands to be replaced in a way that the target bit rate is achieved. In one embodiment, the sequence controller starts with a very soft selection criterion, resulting in a very low number of frequency bands being selected for replacement. If the resulting bit rate after the replacement is still higher than the target bit rate, the sequence controller needs to tighten the selection criterion.

A flowchart of the iteration carried out to achieve the target bit rate is depicted in Fig. 5. The sequence controller 150 checks, in a first verification block 510, whether the target bit rate is achieved. If the target bit rate is not achieved, the sequence controller 150 tightens

the selection criterion in a step 520 and passes the tightened selection criterion onto the identifier 120, upon which new frequency bands for replacement are identified in a block 530 and, finally, the replacer 130 replaces the new  
5 identified frequency bands as well in a step 540. After that, the sequence controller 150 again verifies as to whether the target bit rate has been achieved in a step 510. Once the target bit rate has been achieved, the data can be provided with the target bit rate in a step 550.

10

At the identifier 120, post analyzers can be operative in one embodiment in order to analyze the data according to a selection criterion. The post analyzer operates similar to the pre-analyzer mentioned as being in one embodiment of  
15 the inventive provider 110. Again, analysis-by-synthesis can be carried out by the post analyzer.

Fig. 6 shows a flowchart of an embodiment of a method to carry out analysis-by-synthesis. In a first step 610, an iteration index  $i$  is initialized with value 1. In the  
20 embodiment depicted in Fig. 6, it is assumed that the digital audio data is divided into  $N$  sub-bands. In a step 620, a band is selected according to the iteration index, i.e. the selection process is started with the first  
25 frequency band. In a next step 630, the selected frequency band is replaced with the according noise parameter, upon which in step 640, the entire digital audio data is synthesized together. Once the data is synthesized, a quality criterion or a quality measure can be determined in  
30 step 650. This quality measure can then be stored together with the iteration index indicating the frequency band. In step 660, it is verified whether the iteration has been completed, i.e. if all frequency bands have been checked already and, if not, the iteration index is increased by  
35 one step in step 670 and the next band is selected again in step 620. Once the entire iteration process has been completed, i.e. if all  $N$  frequency bands have been tested, the frequency bands with the lowest quality impact can be

chosen and be identified for replacement. The quality impact can be determined by traditional measures as, for example, a signal-to-noise ratio. Another measure would be a measure that is determined by a psycho-acoustically  
5 model, again determining the lowest quality impact for the human auditory system.

The criterion for noise substitution during the encoding process, as indicated in Fig. 3 at the provider 110 as well  
10 as the selection criterion carried out by the post-analyzer within the identifier 120 can basically refer to the same measure. However, the pre-selection criterion, as it is used in one embodiment at the provider, determines frequency bands within digital audio data, which do not  
15 harm the quality of the digital audio data, which is again determined by the psycho-acoustical model. Deferring from that objective, i.e. decreasing the quality and introducing an impact on the quality of the digital audio data considering the human auditory system, the post-analyzer at  
20 the identifier selects frequency bands. Although the pre-selection criterion and the selection criterion can refer to the same measure, they defer in their impact on the quality.

25 Measures that can be taken at the pre-analyzer as well as the post-analyzer being used as pre-selection criterion or selection criterion are, for example, a lowest tonality, a lowest or highest signal-to-noise ratio, a lowest or highest signal-to-mask ratio, i.e. taking into account the  
30 human auditory system properties, a lowest energy in a frequency band, a highest center frequency of a frequency band or a best stability in the time domain, i.e. lowest variability in a time period.

35 In another embodiment, the replacer 130 is adapted to replace frequency bands, which are consecutive frequency bands together with a single noise synthesis parameter,

i.e. by replacing several frequency band data carrying out a higher bit rate reduction of the digital audio data.

While, in the state of the art, codec perceptual noise substitution is used to replace scaling factors judged to be noise-like before the actual quantization and coding step, noise substitution is used in embodiments of the present invention to reduce the bit rate. There are more useful cases for perceptual noise substitution than just merely replacing scale factor bands found to be noise-like in the perceptual model, as it is currently achieved by the state of the art. In embodiments of the present invention, perceptual noise substitution is employed as part of a constraints reduction apparatus or bit rate reduction apparatus in the more advanced constraints reduction method.

A full flow chart of the state of the art encoding process extended by an embodiment of the inventive method is shown in Fig. 7. Fig. 7 shows the input signal being input into a filter bank 705 and into a perceptual model 710. The frequency coefficients being output from the filter bank 705 are then input into a bit/noise allocation block 715, which is also connected to the perceptual model block 710. The bit/noise allocation block 715 is followed by a quantization block 720 and by an irrelevancy reduction block 725, which are similar to the bit/noise allocation block 820 and the quantization block 830 as explained in Fig. 8. After the irrelevancy reduction block 725, a code requirement verification is done in block 730. If the coding requirement is met, the entropy encoded quantized frequency coefficients and the coded scaling factors are input to a bit stream multiplexer 735 and the encoded data is available with the required bit rate. If the coding requirement, which is verified in the coding requirement block 730 is not met, another verification step is done in 740, which checks as to whether a further reduction of the bit rate is possible without introducing spectral holes. If

it is possible to further reduce the bit rate without introducing spectral holes, the coding demand is reduced in block 745 and the relaxation is limited, so spectral holes cannot be introduced in a following step 750. The process  
5 is then repeated starting with the bit/noise allocation step 715.

This state of the art procedure is extended by an embodiment of an inventive method within the box 755 in  
10 Fig. 7. If, in the verification step 740 it is determined that no further reduction in the bit rate of the digital audio data is possible without introducing spectral holes, the procedure is followed by a selection block 760. The selection block 760 selects the most suitable scale factor  
15 band for artificial noise substitution, also called perceptual noise substitution. Once a proper frequency band has been identified, the perceptual noise is generated in a block 765 inserted into the digital data, where the selected scale factor band is removed from the quantized  
20 spectrum array in step 770 and the coding demand can be recalculated in step 775. After this, the coding requirement can be verified in step 780 and if the coding requirement is not met, it is returned to the step 760, i.e. the next frequency band is selected for perceptual  
25 noise substitution. Eventually, the process will terminate with a coding requirement that is met, upon which the bit stream can be multiplexed in a step 735 and the digital data is available with reduced bit rate.

30 As can be seen from Fig. 7, an embodiment of the present invention is, in the upper-part of the process flow, very similar to an advanced coding solution that can be found in the state of the art described above. The difference lies in the constraint reduction options, where embodiments of  
35 the present invention prevent the introduction of spectral holes. Instead of removing scale factor bands and introducing spectral holes, embodiments of the present invention solve the problem in a more effective way.



Principally, in a first step, selection of the most appropriate scale factor bands, or a sub-set of frequency coefficients, for substitution with artificial noise in the decoder is carried out.

5

This selection can be done by various means, such as one of, or a multiple of, a scale factor band with the lowest tonality, a scale factor band with the lowest or highest signal-to-noise ratio, a scale factor band with the lowest or highest signal-to-mask ratio, a scale factor band with the lowest energy, a scale factor band with the highest center frequency, a scale factor band with the best stability in the time domain or any grouping of frequency coefficients fulfilling one or more of the just mentioned metrics.

10  
15

It is noted that these means are just explanatory and other means known to a person skilled in the art, as they are within the scope and spirit of this invention.

20

After the selection has been carried out, selected scale factor bands or other grouping of frequency coefficients are coded, for example, with the perceptual noise substitution tool, meaning that the embodiments of the present invention remove the spectral content from the digital audio data and instead of the scaling factors for the band, for example, its approximate average energy is transmitted along with an appropriate flag telling the decoder to reconstruct said band with artificially-generated noise of approximately the same energy as transmitted in the bit stream.

25  
30

In another embodiment of the present invention following the perceptual noise substitution coding, the bit demand of the replaced spectral coefficients can now be removed from the quantized spectrum bit demands and the total bit demands can be compared to the encoder constraints. If the constraints are still not met, the procedure continues

35

until constraints are either met or all bands are coded with the perceptual noise substitution. Therefore, it is necessary to set a minimum constraint such that the perceptual noise substitution energy factors could be transmitted for all the bands. If it is desirable to reach such limits, it is possible to employ the removal of the perceptual noise substitution scale factors to reach even very high coding constraints. This could be achieved by iteratively removing most suitable perceptual noise substitution factors, where methods for evaluating such factors are known to a person skilled in the art, for example, like the selection of the lowest energy scale factor or the highest frequency scale factor, etc. The bit demand is then re-evaluated and the process is repeated until it satisfies the constraints or, respectively, all factors are set to zero.

Embodiments of the present invention provide the advantage that the introduction of spectral holes is effectively prevented, as artifacts connected to the spectral band shut downs or spectral holes, in a modern perceptual audio codec are circumventive, yielding a better quality of digital audio data with respect to the human auditory system.

One embodiment of the present invention is an audio coding apparatus based on frequency-based perceptual audio coding with a perceptual model, time-to-frequency mapping and quantization and an entropy coding block. Furthermore, coding can be based on the grouping of a plurality of frequency domain spectral coefficients to scale factor bands and quantizing them with irrelevancy reduction. In another embodiment, the plurality of frequency domain spectral coefficients can be treated in a manner proportional with the critical bands of the human auditory system and quantizing them with irrelevancy reduction. Another embodiment of the present invention comprises the transmission of said coefficients in a coded bit stream.

Moreover, an embodiment could make use of substitution of the scale factor band with the artificially-generated narrow band noise in the decoder without the need to transmit the spectral contents of a said scale factor band, where the coding constraint's evaluation methods can be based on just noticeable distortion measures calculated by a perceptual model and the values of the spectral coefficients. Embodiments of the present invention reduce the coding requirements in order to meet the coding constraints by substitution of the scaling factor bands with one of the methods described above. For example, a suitable scale factor band can be selected for reduction of coding requirements by determining the scale factor band with the most similarity to white noise, the scale factor band with the highest center frequency, the scale factor band with the lowest energy, the scale factor band with the highest signal-to-noise ratio, the scale factor band with the lowest signal-to-noise ratio, the scale factor band with the highest signal to just noticeable distortion energy ratio or the scale factor band with the lowest signal to just noticeable distortion energy ratio.

Depending on certain implementation requirements of the inventive methods, the inventive methods can be implemented in hardware or software. The implementation can be performed using a digital storage medium, in particular a disc, DVD or a CD having an electronically-readable control signal stored thereon, which operates with a programmable computer system, such that the inventive methods are performed. Generally, the present invention is, therefore, a computer program product with a program code stored on a machine-readable carrier, the program code being operative for performing the inventive methods when the computer program product runs on a computer. In other words, the inventive methods are, therefore, a computer program having a program code for performing at least one of the inventive methods when the computer program runs on a computer.

Reference List

- 100 Apparatus for encoding digital audio data  
110 Provider  
5 112 Filter bank  
114 Scale factor quantization and noise substitution  
116 Psycho-acoustic model and pre-analyzer  
120 Identifier  
130 Replacer  
10 140 Entropy encoder  
150 Sequence controller  
510 Target bit rate verification  
520 Selection criterion tightening  
530 Identification of frequency bands  
15 540 Replacement of frequency band data  
550 Provision of data  
610 Initialize i  
620 Select band i  
630 Replace band i  
20 640 Synthesize total digital audio data  
660 All frequency bands verified  
670 Increase iteration index i  
680 Identifier band  
705 Filter bank  
25 710 Perceptual model  
715 Bit/noise allocation  
720 Quantization  
725 Irrelevancy reduction  
730 Code requirement verification  
30 735 Bit stream multiplexer  
740 Verification further bit rate reduction without  
spectral holes  
745 Reduction of coding demand  
750 Limiting of relaxation so spectral holes cannot be  
35 introduced  
755 Embodiment of inventive method  
760 Selection of most suitable band  
765 Perceptual noise substitution

- 770 Removal of selected scale factors of selected frequency bands
- 775 Recalculation of coding demand
- 780 Verification of coding requirements
- 5 810 Filter bank
- 815 Perceptual model
- 820 Bit/noise allocation
- 825 Quantization
- 830 Irrelevancy reduction
- 10 835 Coding verification
- 840 Reduction of bit demand
- 845 Bit stream multiplexer

15

Claims

1. Apparatus for encoding digital audio data with a reduced bit rate, comprising:
- 5 a provider of psycho-acoustically quantized digital audio data with a bit rate being higher than the reduced bit rate;
- 10 identifier for identifying the frequency band according to a selection criterion, the selection criterion being such that an impact on the quality of the digital audio data when the data in the identified frequency band is replaced by generated noise is smaller than the impact
- 15 on the quality of the digital audio data, which would arise when data in a different frequency band is replaced by generated noise, and
- a replacer for replacing data in the identified
- 20 frequency band of the digital audio data by a noise synthesis parameter, the noise synthesis parameter requiring a smaller amount of data than the data in the identified frequency band, the digital audio data having the reduced bit rate.
- 25
2. Apparatus of claim 1, wherein the provider is adapted for providing psycho-acoustically quantized digital audio data per frequency band, the frequency band being determined by a filter in a filter bank.
- 30
3. Apparatus of claim 1, further comprising an entropy encoder for encoding the digital audio data having the reduced bit rate.
- 35 4. Apparatus of one of the claims 1 to 3, wherein the psycho-acoustically encoded digital data includes entropy encoded quantized spectral data and wherein the provider includes an entropy decoder for entropy

decoding the psycho-acoustically encoded digital audio data for providing the psycho-acoustically quantized spectral data and wherein the identifier and the replacer are operative to process the entropy decoded  
5 psycho-acoustically quantized digital audio data.

5. Apparatus of one of the claims 1 to 4, wherein the provider includes a noise substitution process for replacing spectral data in a pre-selected frequency  
10 band by an inserted parameter of the noise substitution process, the pre-selected frequency bands being identified by a pre-selection criterion, the noise substitution process being carried out instead psycho-acoustically quantizing the digital audio data.

15

6. Apparatus of claim 5, wherein the provider includes a pre-analyzer for analyzing digital audio data according to the pre-selection criterion for pre-selecting the frequency band for insertion of a noise substitution  
20 parameter.

20

7. Apparatus of one of the claims 1 to 6, wherein the identifier includes a post-analyzer for analyzing psycho-acoustically quantized data in a frequency band according to the selection criterion for identifying the frequency band for psycho-acoustically quantized data replacements.  
25

25

8. Apparatus of one of the claims 5 to 7, wherein the pre-analyzer or the post-analyzer are operative to utilize the pre-selection criterion or the selection criterion, the pre-selection criterion being different from the selection criterion, the pre-selected frequency band being different from the identified frequency bands.  
30

30

35

9. Apparatus of claim 8, wherein the pre-analyzer utilizes the pre-selection criterion and the post-analyzer utilizes the selection criterion corresponding to one

- of or a combination of the group of a lowest tonality, a lowest or highest signal-to-noise ratio, a lowest or highest signal-to-mask ratio, a lowest energy, a highest central frequency; a best stability in the time domain or a lowest variability in the time domain.
- 5
10. Apparatus of one of the claims 1 to 9, further comprising a sequence controller for controlling the identifier and the replacer, the sequence controller comparing the reduced bit rate with a target bit rate, adapting the selection criterion such that more frequency bands are identified for replacement by noise synthesis parameters when the reduced bit rate is higher than the target bit rate.
- 10
- 15
11. Apparatus of one of the claims 1 to 10, wherein the replacer is adapted for replacing data of a plurality of frequency bands and for replacing data of consecutive frequency bands by a noise synthesis parameter.
- 20
12. Apparatus of one of the claims 1 to 11, wherein the provider is operative to provide psycho-acoustically quantized data from encoded digital audio data, the encoded digital audio data being encoded according to ISO/IEC 14496.
- 25
13. Apparatus of one of the claims 3 to 12, being adapted for encoding digital audio data with reduced bit rate according to ISO/IEC 14496.
- 30
14. A method for encoding digital audio data with a reduced bit rate, comprising the steps of:
- 35
- providing psycho-acoustically quantized digital audio data with a bit rate being higher than the reduced bit rate;



identifying the frequency band according to a selection criterion, the selection criterion being such that an impact on a quality of the digital audio data when the data in the identified band replaced by generated noise is smaller than the impact on the quality of the digital audio data, which would arise when data in a different frequency band is replaced by generated noise, and

replacing data in the identified frequency band of the digital audio data with noise synthesis parameter, the noise synthesis parameter requiring a smaller amount of data than the data in the identified frequency band, the digital audio data having the reduced bit rate.

15

15. A computer program having program codes for performing the method of claim 14 when the program code runs in a computer.

FIG 1

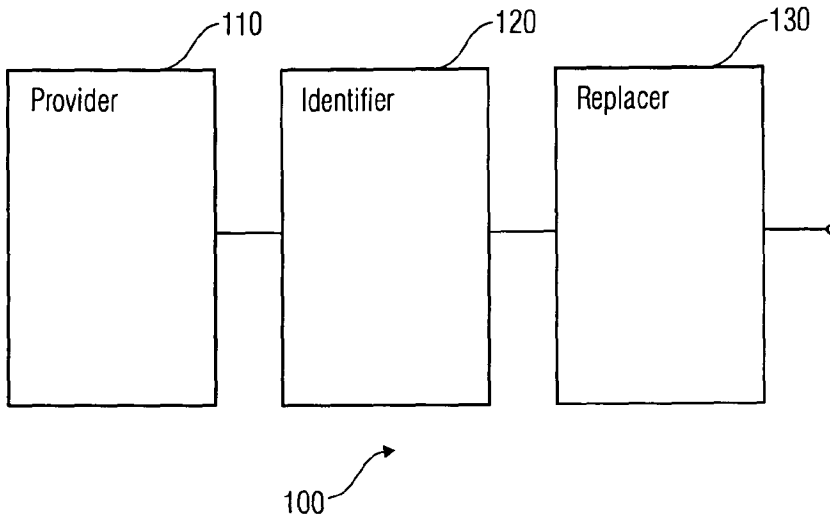


FIG 2

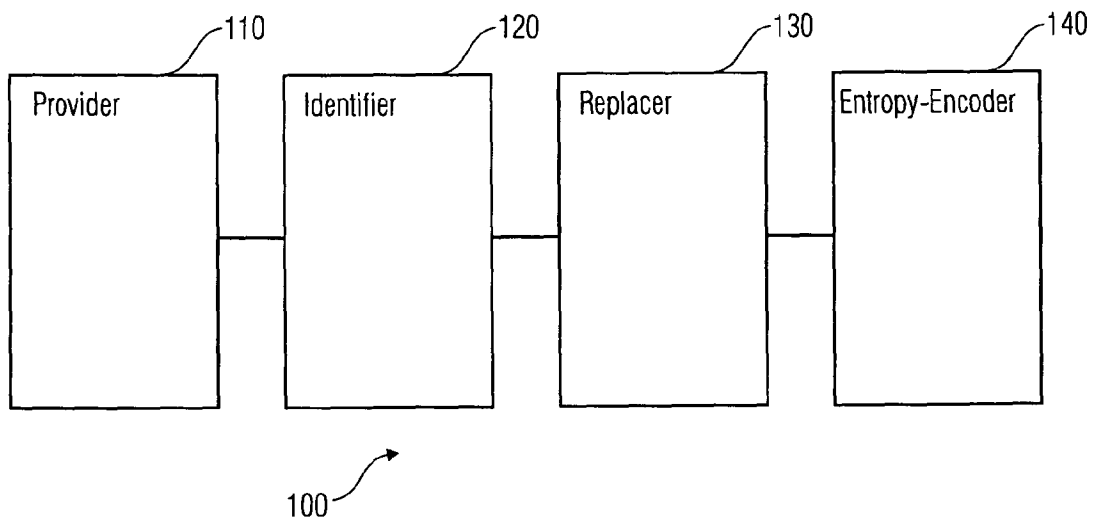


FIG 3

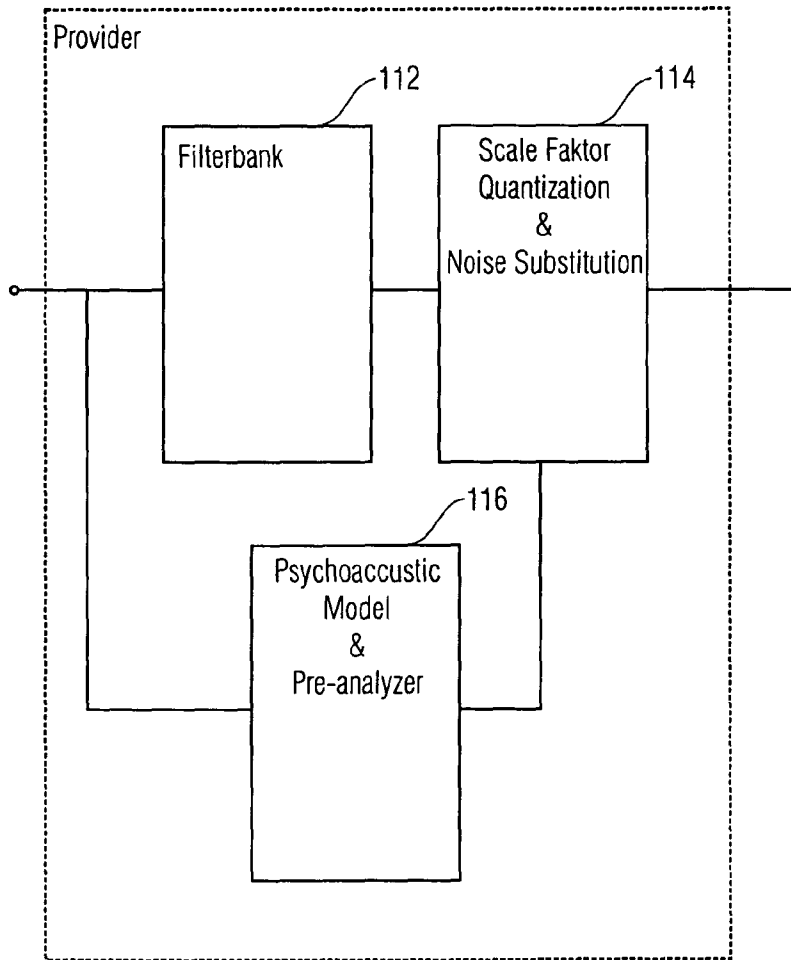


FIG 4

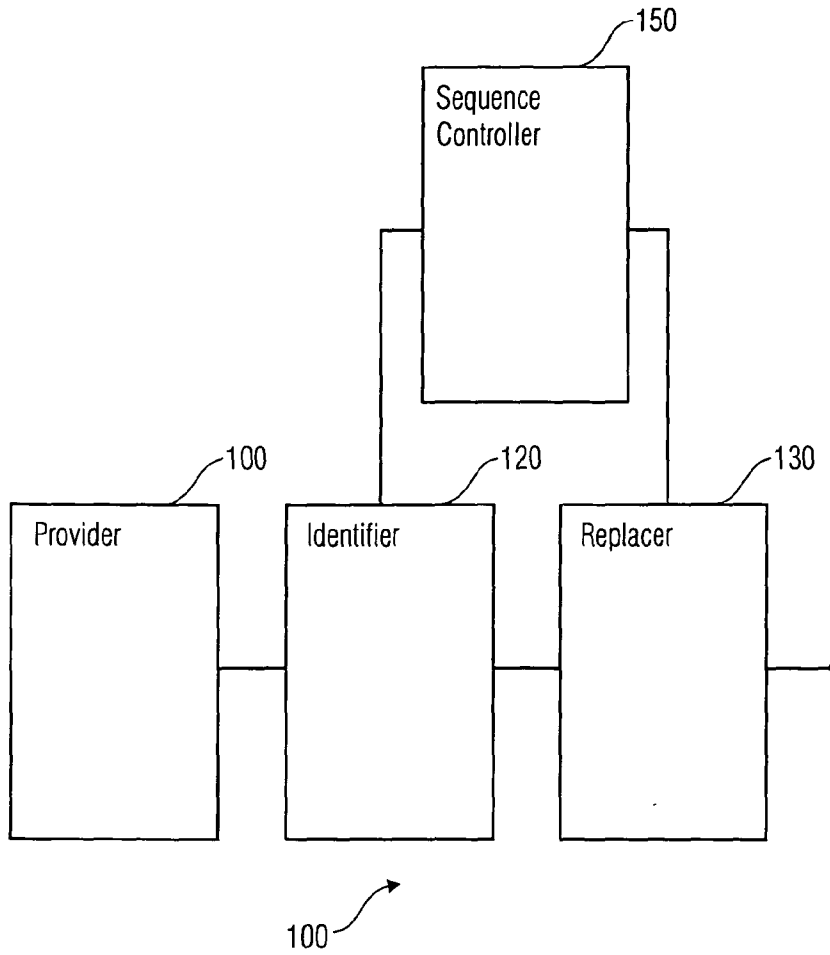


FIG 5

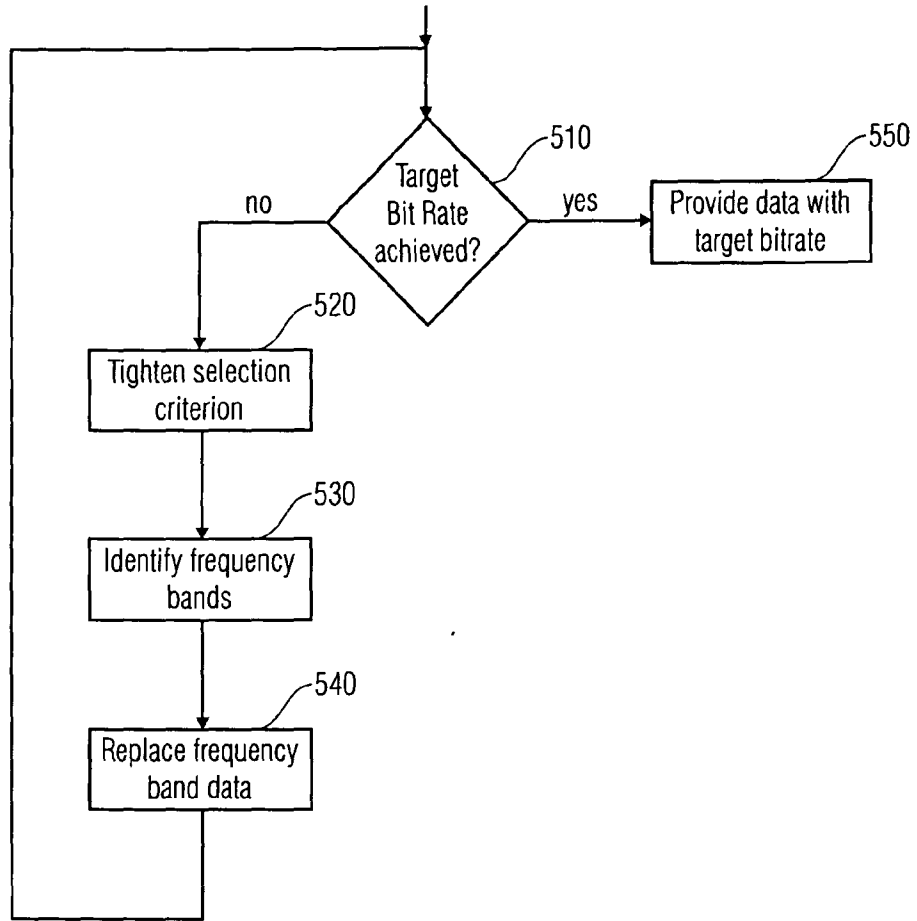


FIG 6

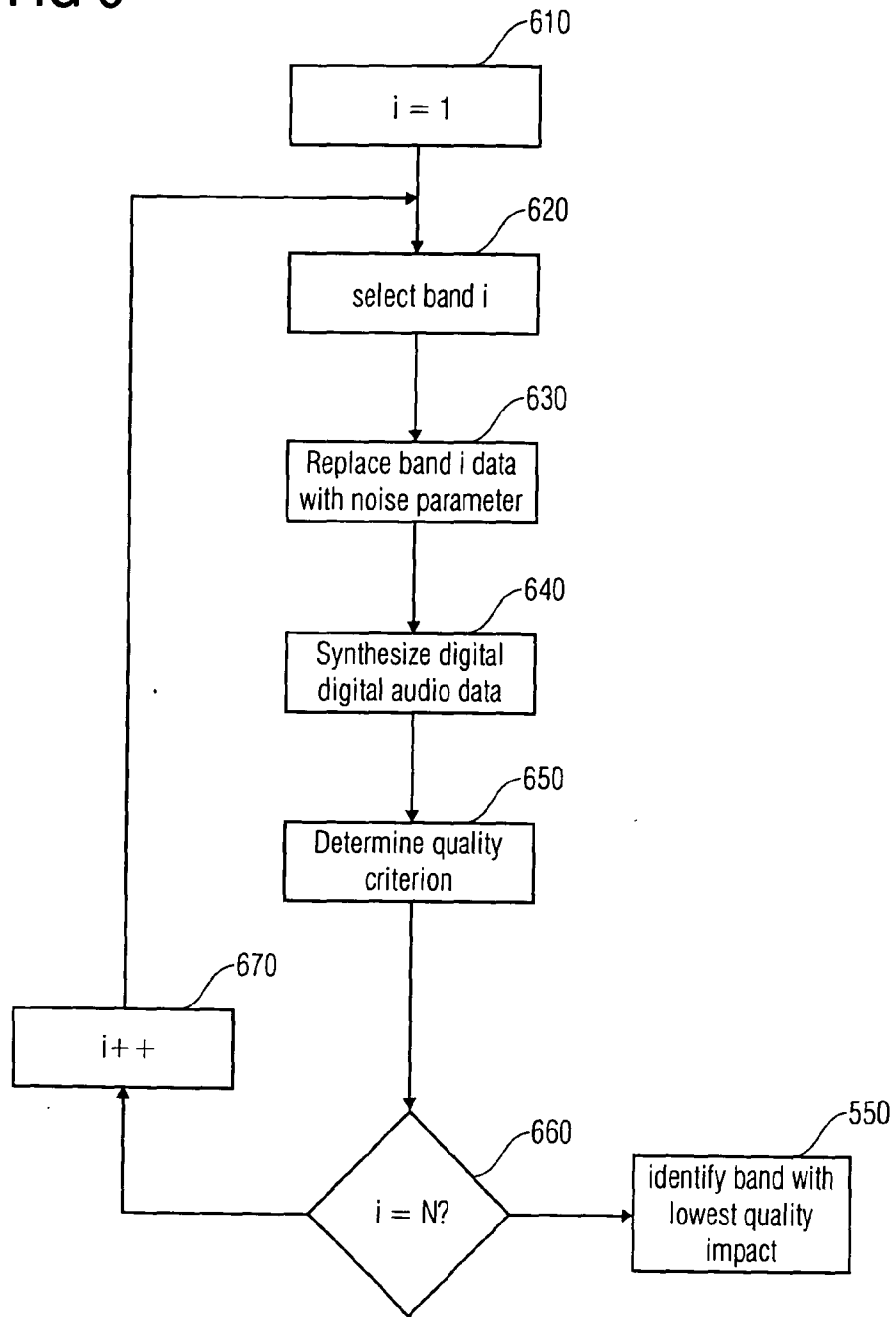


FIG 7

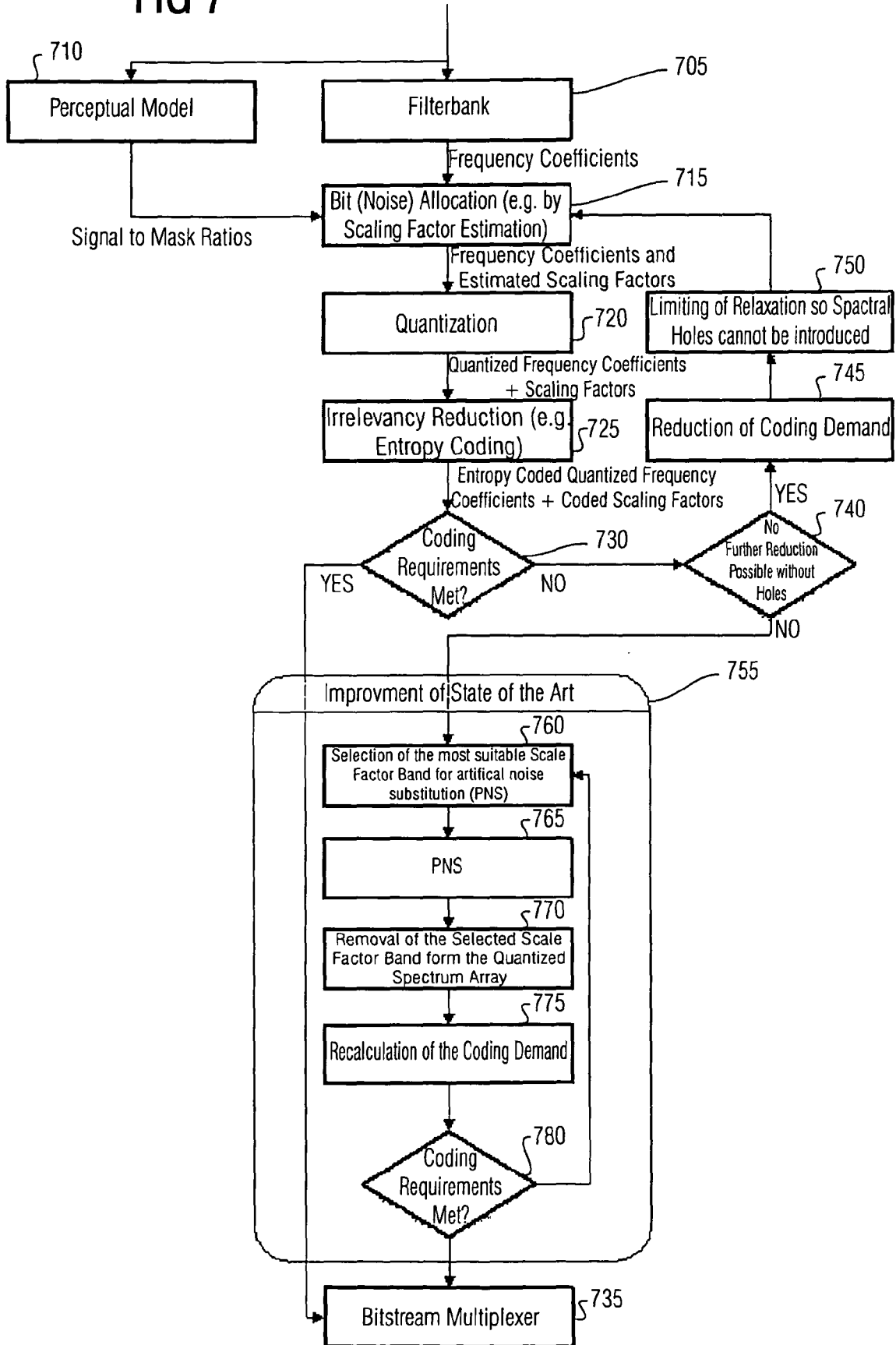


FIG 8

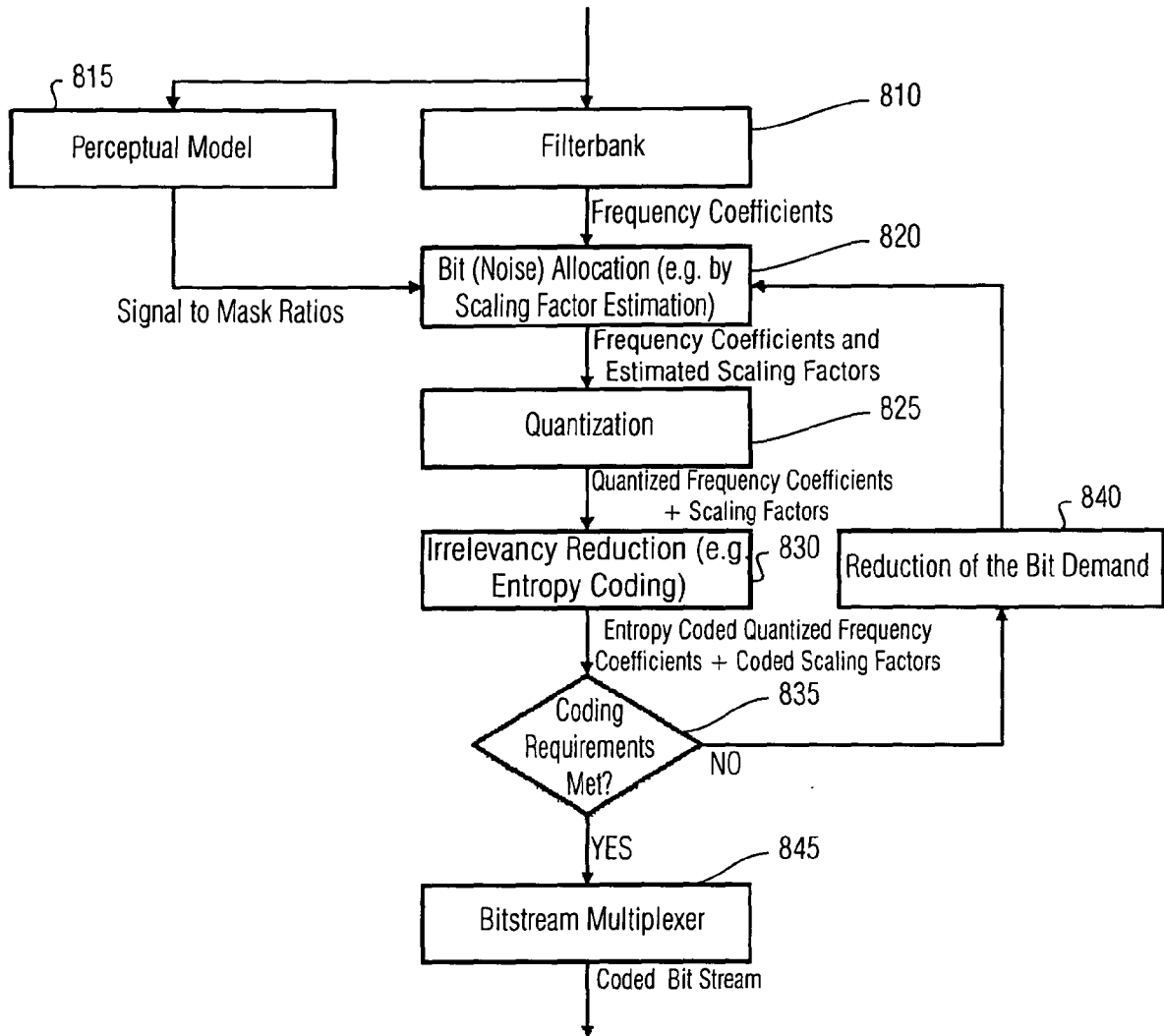
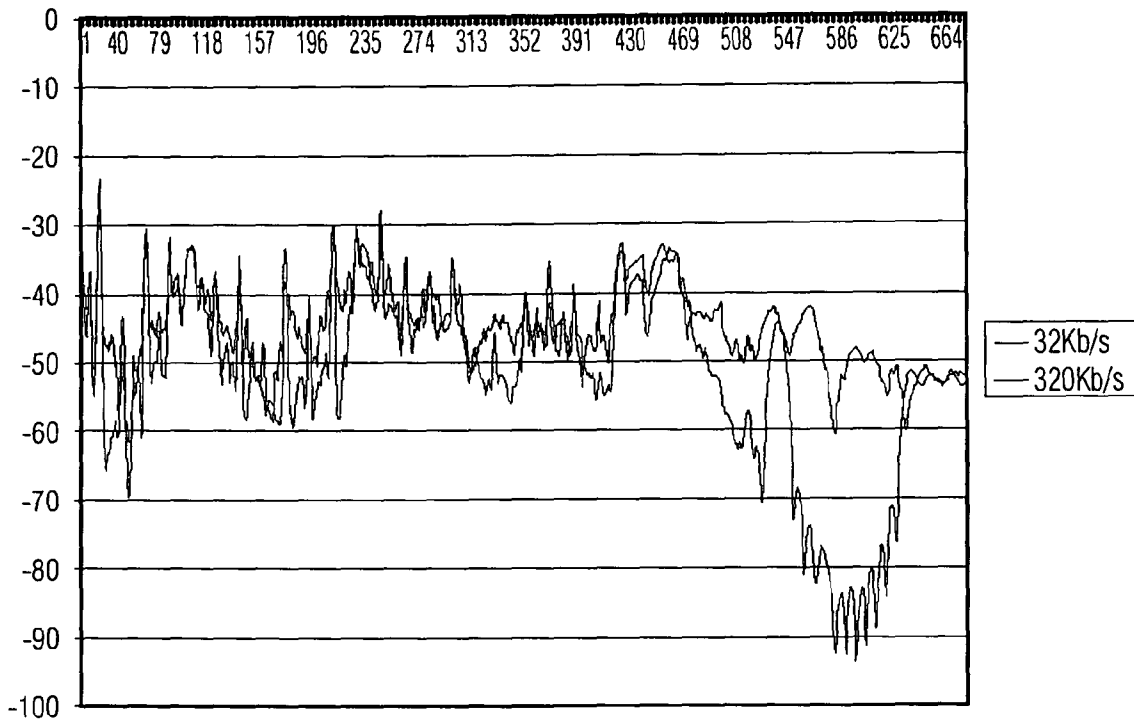




FIG 9



**INTERNATIONAL SEARCH REPORT**

International application No  
PCT/EP2006/009601

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. G10L19/02				
According to International Patent Classification (IPC) or to both national classification and IPC				
<b>B. FIELDS SEARCHED</b>				
Minimum documentation searched (classification system followed by classification symbols) 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, WPI Data, INSPEC, IBM-TDB				
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
A	"3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General audio codec audio processing functions; Enhanced acPlus general audio codec; Encoder specification AAC part (Release 6); 3GPP TS 26.403 V6.0.0" 3GPP, September 2004 (2004-09), pages 1-23, XP002410983 Sophia Antipolis (FR) page 14, paragraph [5.6.1] page 18, paragraph [5.6.1.3.3]  ----- -/--	1-15		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <span style="margin-left: 200px;"><input checked="" type="checkbox"/> See patent family annex.</span>				
* Special categories of cited documents : <table style="width:100%; border:none;"> <tr> <td style="width:50%; border:none;">                         "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                     </td> <td style="width:50%; border:none;">                         "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.                          "&amp;" document member of the same patent family                     </td> </tr> </table>			"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
"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  11 December 2006		Date of mailing of the international search report  15/01/2007		
Name and mailing address of the ISA/ 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		Authorized officer  Greiser, Norbert		

INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2006/009601

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, 1998, pages 1-14, XP008006769 abstract page 2, paragraph 2 - page 3, paragraph 3 -----	1
A	US 2003/233234 A1 (TRUMAN MICHAEL MEAD [US] ET AL) 18 December 2003 (2003-12-18) abstract page 2, paragraphs [0010] - [0011] page 5, paragraphs [0056] - [0063] figures 3-5 -----	1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2006/009601

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2003233234 A1	18-12-2003	AU 2003237295 A1	31-12-2003
		CA 2489441 A1	24-12-2003
		CN 1662958 A	31-08-2005
		EP 1514261 A1	16-03-2005
		JP 2005530205 T	06-10-2005
		MX PA04012539 A	28-04-2005
		WO 03107328 A1	24-12-2003
		US 2003233236 A1	18-12-2003

---