



(22) Date de dépôt/Filing Date: 2012/02/13

(41) Mise à la disp. pub./Open to Public Insp.: 2012/08/23

(45) Date de délivrance/Issue Date: 2017/08/29

(62) Demande originale/Original Application: 2 827 266

(30) Priorité/Priority: 2011/02/14 (US61/442,632)

(51) Cl.Int./Int.Cl. *G10L 19/24* (2013.01),
G10L 19/02 (2013.01), *G10L 19/025* (2013.01),
G10L 19/12 (2013.01)

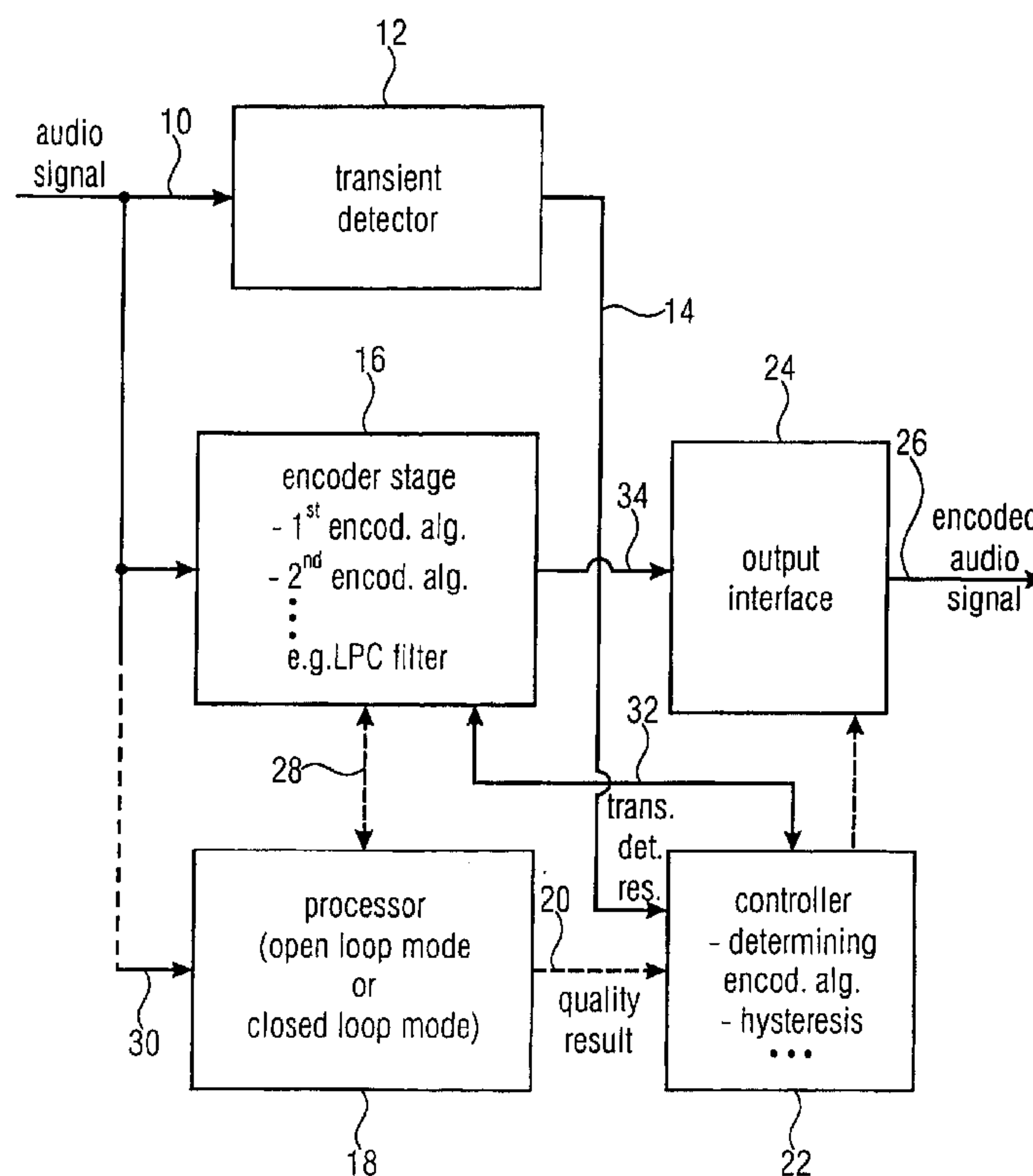
(72) Inventeurs/Inventors:
HELMRICH, CHRISTIAN, DE;
FUCHS, GUILLAUME, DE;
MARKOVIC, GORAN, DE

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

(74) Agent: BORDEN LADNER GERVAIS LLP

(54) Titre : APPAREIL ET PROCEDE DE CODAGE D'UNE PARTIE D'UN SIGNAL AUDIO AU MOYEN D'UNE
DETECTION DE TRANSITOIRE ET D'UN RESULTAT DE QUALITE

(54) Title: APPARATUS AND METHOD FOR CODING A PORTION OF AN AUDIO SIGNAL USING A TRANSIENT
DETECTION AND A QUALITY RESULT



(57) Abrégé/Abstract:

An apparatus for coding a portion of an audio signal (10) to obtain an encoded audio signal (26) for the portion of the audio signal comprises a transient detector (12) for detecting whether a transient signal is located in the portion of the audio signal to obtain a

(57) **Abrégé(suite)/Abstract(continued):**

transient detection result (14), an encoder stage (16) for performing a first encoding algorithm on the audio signal, the first encoding algorithm having a first characteristic, and for performing a second encoding algorithm on the audio signal, the second encoding algorithm having a second characteristic being different from the first characteristic, a processor (18) for determining which encoding algorithm results in an encoded audio signal being a better approximation to the portion of the audio signal with respect to the other encoding algorithm to obtain a quality result (20), and a controller (22) for determining whether the encoded audio signal for the portion of the audio signal is to be generated by either the first encoding algorithm or the second encoding algorithm based on the transient detection result (14) and the quality result (20).

Abstract

An apparatus for coding a portion of an audio signal (10) to obtain an encoded audio signal (26) for the portion of the audio signal comprises a transient detector (12) for detecting whether a transient signal is located in the portion of the audio signal to obtain a transient detection result (14), an encoder stage (16) for performing a first encoding algorithm on the audio signal, the first encoding algorithm having a first characteristic, and for performing a second encoding algorithm on the audio signal, the second encoding algorithm having a second characteristic being different from the first characteristic, a processor (18) for determining which encoding algorithm results in an encoded audio signal being a better approximation to the portion of the audio signal with respect to the other encoding algorithm to obtain a quality result (20), and a controller (22) for determining whether the encoded audio signal for the portion of the audio signal is to be generated by either the first encoding algorithm or the second encoding algorithm based on the transient detection result (14) and the quality result (20).

Apparatus and Method for Coding a Portion of an Audio Signal Using a Transient Detection and a Quality Result

5

Specification

The present invention is related to audio coding and, particularly, to switched audio coding, where, for different time portions, the encoded signal is generated using different
10 encoding algorithms.

Switched audio coders which determine different encoding algorithms for different portions of the audio signal are known. An example is the so-called extended adaptive multi-rate-wideband codec or AMR-WB+ codec defined in the International Standard
15 3GPP TS 26.290 V6.1.0 2004-12. In this technical specification, the coding concept is described, which extends the ACELP (Algebraic Code Excited Linear Prediction) based AMR-WB codec by adding TCX (Transform Coded Excitation), bandwidth extension, and stereo. The AMR-WB+ audio codec processes input frames equal to 2048 samples at an internal sampling frequency F_s . The internal sampling frequency is limited to the range
20 12,800 to 38,400 Hz. The 2048 sample frames are split into two critically sampled equal frequency bands. This results in two superframes of 1024 samples corresponding to the low-frequency (LF) and high-frequency (HF) bands. Each superframe is divided into four 256-samples frames. Sampling at the internal sampling rate is obtained by using a variable sampling conversion scheme, which re-samples the input signal. The LF and HF signals
25 are then encoded using two different approaches. The LF signal is encoded and decoded using the "core" encoder/decoder, based on switched ACELP and TCX. In the ACELP mode, the standard AMR-WB codec is used. The HF signal is encoded with relatively few bits (16 bits/frame) using a bandwidth extension (BWE) method.

30 The parameters transmitted from encoder to decoder are the mode-selection bits, the LF parameters and HF signal parameters. The parameters for each 1024-sample superframe are decomposed into four packets of identical size. When the input signal is stereo, the left and right channels are combined into mono-signals for a ACELP-TCX encoding, whereas the stereo encoding receives both input channels. In the AMR-WB+ decoder structure, the
35 LF and HF bands are decoded separately. Then, the bands are combined in a synthesis filterbank. If the output is restricted to mono only, the stereo parameters are omitted and the decoder operates in mono mode.

The AMR-WB+ codec applies LP (Linear Prediction) analysis for both the ACELP and TCX modes, when encoding the LF signal. The LP coefficients are interpolated linearly at every 64-sample sub-frame. The LP analysis window is a half-cosine of length 384 samples. The coding mode is selected based on closed-loop analysis-by-synthesis method.

5 Only 256 sample frames are considered for ACELP frames, whereas frames of 256, 512 or 1024 samples are possible in TCX mode. The ACELP coding consists of long-term prediction (LTP) analysis and synthesis and algebraic codebook excitation. In the TCX mode, a perceptually weighted signal is processed in the transform domain. The Fourier transformed weighted signal is quantized using split multi-weight lattice quantization

10 (algebraic vector quantization). The transform is calculated in 1024, 512 or 256 sample windows. The excitation signal is recovered by inverse filtering a quantized weighted signal through the inverse weighting filter. In order to determine whether a certain portion of the audio signal is to be encoded using the ACELP mode or the TCX mode, a closed-loop mode selection or an open-loop mode selection is used. In a closed-loop mode

15 selection, 11 successive trials are used. Subsequent to a trial, a mode selection is made between two modes to be compared. The selection criterion is the average segmental SNR (Signal Noise Ratio) between the weighted audio signal and the synthesized weighted audio signal. Hence, the encoder performs a complete encoding in both encoding algorithms, a complete decoding in accordance with both encoding algorithms and,

20 subsequently, the results of both encoding/decoding operations are compared to the original signal. Hence, for each encoding algorithm, i.e., ACELP on the one hand and TCX on the other hand, a segmental SNR value is obtained and the encoding algorithm having the better segmental SNR value or having a better average segmental SNR value determined over a frame by averaging over the segmental SNR values for the individual

25 sub-frames is used.

An additional switched audio coding scheme is the so-called USAC coder (USAC = Unified Speech Audio Coding). This coding algorithm is described in ISO/IEC 23003-3. The general structure can be described as follows. First, there is a common pre/post

30 processing system of an MPEG Surround functional unit to handle stereo or multi-channel processing and an enhanced SBR unit generating the parametric representation of the higher audio frequencies of the input signal. Then, there are two branches, one consisting of a modified advanced audio coding (AAC) tool path and the other consisting of a linear prediction coding (LP or LPC domain) based path, which in turn features either a

35 frequency-domain representation or a time-domain representation of the LPC residual. All transmitted spectra for both, AAC and LPC, are represented in MDCT domain following quantization and arithmetic coding. The time-domain representation uses an ACELP excitation coding scheme. The functions of the decoder are to find the description of the

quantized audio spectra or time-domain representation in the bitstream payload and to decode the quantized values and other reconstruction information. Hence, the encoder performs two decisions. The first decision is to perform a signal classification for frequency domain versus linear prediction domain mode decision. The second decision is to determine, within the linear prediction domain (LPD), whether a signal portion is to be encoded using ACELP or TCX.

For applying a switched audio coding scheme in scenarios, where a very low delay is necessary, particular attention has to be paid to transform-based coding parts, since these coding parts introduce a certain delay which depends on the transform length and window design. Therefore, the USAC coding concept is not suitable to very low-delay applications due to the modified AAC coding branch having a considerable transform length and length adaptation (also known as block switching) involving transitional windows.

On the other hand, the AMR-WB+ coding concept was found to be problematic due to the encoder-side decision whether ACELP or TCX is to be used. ACELP provides a good coding gain, but may result in significant audio quality problems when a signal portion is not suitable for the ACELP coding mode. Hence, for quality reasons, one might be inclined to use TCX whenever the input signal does not contain speech. However, using TCX too much at low bitrates will result in bitrate problems, since TCX provides a relatively low coding gain. When one, therefore, looks more onto the coding gain, one might use ACELP whenever possible, but, as stated before, this can result in audio quality problems due to the fact that ACELP is not optimal, for example, for music and similar stationary signals.

The segmental SNR calculation is a quality measure, which determines the better coding mode only based on the result, i.e., whether the SNR between the original signal or the encoded/decoded signal is better, so that the encoding algorithm resulting in a better SNR is used. This, however, always has to operate under bitrate constraints. Therefore, it has been found that only using a quality measure such as, for example, the segmental SNR measure does not always result in the best compromise between quality and bitrate.

It is the object of the present invention to provide an improved concept for coding a portion of an audio signal.

This object is achieved by an apparatus for coding a portion of an audio signal.

3a

According to one aspect of the invention, there is provided an apparatus for coding a portion of an audio signal to obtain an encoded audio signal for the portion of the audio signal, comprising a transient detector for detecting whether a transient signal is located in the portion of the audio signal to obtain a transient detection result; an encoder stage for performing a first encoding algorithm on the audio signal, the first encoding algorithm having a first characteristic, and for performing a second encoding algorithm on the audio signal, the second encoding algorithm having a second characteristic being different from the first characteristic; a processor for determining which encoding algorithm results in an encoded audio signal being a better approximation to the portion of the audio signal with respect to the other encoding algorithm to obtain a quality result; and a controller for determining whether the encoded audio signal for the portion of the audio signal is to be generated by either the first encoding algorithm or the second encoding algorithm based on the transient detection result and the quality result, wherein the controller is configured for applying a hysteresis processing so that the second encoding algorithm or the first encoding algorithm is only determined when the lower quality result indicates a lower quality for the second encoding algorithm or the first encoding algorithm, when a number of earlier signal portions having the first encoding algorithm or the second encoding algorithm, respectively, is equal or lower than a predetermined number, and when the transient detection result indicates a predefined state of the two possible states comprising non-transients and transients.

According to another aspect of the invention, there is provided a method of coding a portion of an audio signal to obtain an encoded audio signal for the portion of the audio signal, comprising detecting whether a transient signal is located in the portion of the audio signal to obtain a transient detection result; performing a first encoding algorithm on the audio signal, the first encoding algorithm having a first characteristic, and performing a second encoding algorithm on the audio signal, the second encoding algorithm having a second characteristic being different from the first characteristic; determining which encoding algorithm results in an encoded audio signal being a better approximation to the portion of the audio signal with respect to the other encoding algorithm to obtain a quality result; and determining whether the encoded audio signal for the portion of the audio signal is to be generated by either the first encoding algorithm or the second encoding algorithm based on the transient detection result and the quality result, wherein the determining comprises applying a hysteresis processing

3b

so that the second encoding algorithm or the first encoding algorithm is only determined when the lower quality result indicates a lower quality for the second encoding algorithm or the first encoding algorithm, when a number of earlier signal portions having the first encoding algorithm or the second encoding algorithm, respectively, is equal or lower than a predetermined number, and when the transient detection result indicates a predefined state of the two possible states comprising non-transients and transients.

The present invention is based on the finding that a better decision between a first encoding algorithm suited for more transient signal portions and a second encoding algorithm suitable for more stationary signal portions can be obtained when the decision is not only based on a quality measure but, additionally, on a transient detection result. While the quality measure only looks at the result of the encoding/decoding chain with respect to the original signal, the transient detection result additionally relies on an analysis of the original input audio signal alone. Hence, it has been found out that a combination of both measures, i.e., the quality result on the one hand and the transient detection result on the other hand for finally determining whether a portion of an audio signal is to be encoded by which encoding algorithm leads to an improved compromise between coding gain on the one hand and audio quality on the other hand.

An apparatus for coding a portion of an audio signal to obtain an encoded audio signal for the portion of an audio signal comprises a transient detector for detecting whether a transient signal is located in the portion of the audio signal to obtain a transient detection result. The apparatus furthermore comprises an encoder stage for performing a first encoding algorithm on the audio signal, the first encoding algorithm having a first characteristic, and for performing a second encoding algorithm on the audio signal, the second encoding algorithm having a second characteristic being different from the first characteristic. In an embodiment, the first characteristic associated with the first encoding algorithm is better suited for a more transient signal, and the second encoding characteristic associated with the second encoding algorithm is better suited for more stationary audio signals. Exemplarily, the first encoding algorithm is an ACELP encoding algorithm and the second encoding algorithm is a TCX encoding algorithm which may be based on a modified discrete cosine transform, an FFT transform or any other transform or filterbank. Furthermore, a processor is provided for determining, which encoding algorithm results in an encoded audio signal being a better approximation to the portion of the audio signal to obtain a quality result. Furthermore, a controller is provided, where the controller is configured for determining whether the encoded audio signal for the portion of the audio signal is generated by either the first encoding algorithm or the second encoding algorithm. In accordance with the invention, the controller is configured for performing this determination not only based on the quality result but, additionally, on the transient detection result.

In an embodiment, the controller is configured for determining the second encoding algorithm, although the quality result indicates a better quality for the first encoding algorithm, when the transient detection result indicates a non-transient signal. Furthermore, the controller is configured for determining the first encoding algorithm, although the

quality result indicates a better quality for the second encoding algorithm, when the transient detection result indicates a transient signal.

5 In a further embodiment, this determination, in which the transient result can negate the quality result, is enhanced using a hysteresis function such that the second encoding algorithm is only determined when a number of earlier signal portions, for which the first encoding algorithm has been determined, is smaller than a predetermined number. Analogously, the controller is configured to only determine the first encoding algorithm when a number of earlier signal portions, for which the second encoding algorithm has
10 been determined in the past, is smaller than a predetermined number. An advantage from the hysteresis processing is that the number of switch-overs between coding modes is reduced for certain input signals. A too frequent switch-over at critical points in the signal may generate audible artifacts specifically for low bitrates. The probability of such artifacts is reduced by implementing the hysteresis.

15 In a further embodiment, the quality result is favored with respect to the transient detection result when the quality result indicates a strong quality advantage for one coding algorithm. Then, the encoding algorithm having the much better quality result than the other encoding algorithm is selected irrespective of whether the signal is a transient signal
20 or not. On the other hand, the transient detection result can become decisive when the quality difference between both encoding algorithms is not so high. To this end, it is preferred to not only determine a binary quality result, but a quantitative quality result. A binary quality result would only indicate which encoding algorithm results in a better quality, whereas a quantitative quality result not only determines which encoding
25 algorithm results in a better quality, but how much better the corresponding encoding algorithm is. On the other hand, one could also use a quantitative transient detection result but, basically, a binary transient detection result would be sufficient as well.

30 Hence, the present invention provides a particular advantage with respect to a good compromise between bitrate on the one hand and quality on the other hand, since, for transient signals, the coding algorithm resulting in less quality is selected. When the quality result favors e.g. a TCX decision, nevertheless the ACELP mode is taken, which might result in a slightly reduced audio quality but, in the end, results in a higher coding gain associated with using the ACELP mode.

35 When, on the other hand, the quality result favors an ACELP frame, a TCX decision is, nevertheless, taken for non-transient signals. Hence, the slightly less coding gain is accepted in favor of a better audio quality.

Thus, the present invention results in an improved compromise between quality and bitrate due to the fact that not only the quality of the encoded and again decoded signal is considered but, in addition, also the actually to be encoded input signal is analyzed with
 5 respect to its transient characteristic and the result of this transient analysis is used to additionally influence the decision for an algorithm better suited for transient signals or an algorithm better suited for stationary signals.

Further embodiments of the present invention are subsequently illustrated by reference to
 10 the accompanying drawings, in which:

- Fig. 1 illustrates a block diagram of an apparatus for coding a portion of an audio signal in accordance with an embodiment;
- 15 Fig. 2 illustrates a table for two different encoding algorithms and the signals for which they are suited;
- Fig. 3 illustrates an overview over the quality condition, the transient condition and the hysteresis condition, which can be applied independently of each other, but which are, preferably, applied jointly;
 20
- Fig. 4 illustrates a state table indicating whether a switch-over is performed or not for different situations;
- 25 Fig. 5 illustrates a flowchart for determining the transient result in an embodiment;
- Fig. 6a illustrates a flowchart for determining the quality result in an embodiment;
- Fig. 6b illustrates more details on the quality result of Fig. 6a; and
 30
- Fig. 7 illustrates a more detailed block diagram of an apparatus for coding in accordance with an embodiment.

Fig. 1 illustrates an apparatus for coding a portion of an audio signal provided at an input
 35 line 10. The portion of the audio signal is input into a transient detector 12 for detecting whether a transient signal is located in the portion of the audio signal to obtain a transient detection result on line 14. Furthermore, an encoder stage 16 is provided where the encoder stage is configured for performing a first encoding algorithm on the audio signal, the first

encoding algorithm having a first characteristic. Furthermore, the encoder stage 16 is configured for performing a second encoding algorithm on the audio signal, wherein the second encoding algorithm has a second characteristic which is different from the first characteristic.

5

Additionally, the apparatus comprises a processor 18 for determining which encoding algorithm of the first and second encoding algorithms results in an encoded audio signal being a better approximation to the portion of the original audio signal. The processor 18 generates a quality result based on this determination on line 20. The quality result on line 10 20 and the transient detection result on line 14 are both provided to a controller 22. The controller 22 is configured for determining whether the encoded audio signal for the portion of the audio signal is generated by either the first encoding algorithm or the second encoding algorithm. For this determination, not only the quality result 20, but also the transient detection result 14 are used. Furthermore, an output interface 24 is optionally 15 provided where the output interface outputs an encoded audio signal as, for example, a bitstream or a different representation of an encoded signal on line 26.

In an implementation, where the encoder stage 16 performs an analysis by synthesis processing, the encoder stage 16 receives the same portion of the audio signal and encodes 20 a portion of this audio signal by the first encoding algorithm to obtain the first encoded representation of the portion of the audio signal. Furthermore, the encoder stage generates an encoded representation of the same portion of the audio signal using the second encoding algorithm. Furthermore, the encoder stage 16 comprises, in this analysis by synthesis processing, decoders for both the first encoding algorithm and the second 25 encoding algorithm. One corresponding decoder decodes the first encoded representation using a decoding algorithm associated with the first encoding algorithm. Furthermore, a decoder for performing a further decoding algorithm associated with the second encoding algorithm is provided so that, in the end, the encoder stage not only has the two encoded representations for the same portion of the audio signal, but also the two decoded signals 30 for the same portion of the original audio signal on line 10. These two decoded signals are then provided to the processor via line 28 and the processor compares both decoded representations with the same portion of original audio signal obtained via input 30. Then, a segmental SNR for each encoding algorithm is determined. This so-called quality result provides, in an embodiment, not only an indication of the better coding algorithm, i.e., a 35 binary signal whether the first encoding algorithm or the second encoding algorithm has resulted in a better SNR. Additionally, the quality result indicates a quantitative information, i.e., how much better, for example in dB, the corresponding encoding algorithm is.

In this situation, the controller, when fully relying on the quality result 20, accesses the encoder stage via line 32 so that the encoder stage forwards the already stored encoded representation of the corresponding encoding algorithm to the output interface 24 so that
5 this encoded representation represents the corresponding portion of the original audio signal in the encoded audio signal.

Alternatively, when the processor 18 performs an open-loop mode for determining the quality result, it is not necessary that both encoding algorithms are applied to one and the
10 same audio signal portion. Instead, the processor 18 determines which encoding algorithm is better and, then, the encoder stage 16 is controlled via line 28 to only apply the encoding algorithm indicated by the processor and, then, this encoded representation resulting from the selected encoding algorithm is provided to the output interface 24 via line 34.

15 Depending on the specific implementation of the encoder stage 16, both encoding algorithms may operate in the LPC domain. In this case, such as for ACELP as the first encoding algorithm and TCX as the second encoding algorithm, a common LPC pre-processing is performed. This LPC pre-processing may comprise an LPC analysis of the portion of the audio signal, which determines the LPC coefficients for the portion of the
20 audio signal. Then, an LPC analysis filter is adjusted using the determined LPC coefficients, and the original audio signal is filtered by this LPC analysis filter. Then, the encoder stage calculates a sample-wise difference between the output of the LPC analysis filter and the audio input signal in order to calculate the LPC residual signal which is then subjected to the first encoding algorithm or the second encoding algorithm in an open-loop
25 mode or which is provided to both encoding algorithms in a closed-loop mode as described before. Alternatively, the filtering by the LPC filter and the sample-wise determination of the residual signal can be replaced by the FDNS (frequency domain noise shaping) technology described in the USAC standard.

30 Fig. 2 illustrates a preferred implementation of the encoder stage. As the first encoding algorithm, the ACELP encoding algorithm having an CELP encoding characteristic is used. Furthermore, this encoding algorithm is better suited for transient signals. The second encoding algorithm has a coding characteristic which makes this second encoding algorithm better suited for non-transient signals. Exemplarily, a transform excitation
35 coding algorithm such as TCX is used and, particularly, a TCX 20 encoding algorithm is preferred which has a frame length of 20 ms (the window length can be higher due to an overlap) which makes the coding concept illustrated in Fig. 1 particularly suitable for low-delay implementations which are required in real-time scenarios such as scenarios where

there is a two-way communication as in telephone applications and, particularly, in mobile or cellular telephone applications.

However, the present invention is additionally useful in other combinations of first and second encoding algorithms. Exemplarily, the first encoding algorithm better suited for transient signals may comprise any of well-known time-domain encoders such as GSM-used encoders (G.729) or any other time-domain encoders. The non-transient signal encoding algorithm, on the other hand, can be any well-known transform-domain encoder such as MP3, AAC, AC3 or any other transform or filterbank-based audio encoding algorithm. For a low-delay implementation, however, the combination of ACELP on the one hand and TCX on the other hand, wherein, particularly, the TCX encoder can be based on an FFT or even more preferably on an MDCT with a short window length is preferred. Hence, both encoding algorithms operate in the LPC domain obtained by transforming the audio signal into the LPC domain using an LPC analysis filter. However, the ACELP then operates in the LPC-“time”-domain, while the TCX encoder operates in the LPC-“frequency”-domain.

Subsequently, a preferred implementation of the controller 22 of Fig. 1 is discussed in the context of Fig. 3.

Preferably, the switchover between the first encoding algorithm such as ACELP and the second encoding algorithm such as TCX 20 is performed using three conditions. The first condition is the quality condition represented by the quality result 20 of Fig. 1. The second condition is the transient condition represented by the transient detection result on line 14 of Fig. 1. The third condition is a hysteresis condition which relies on the decisions made by the controller 22 in the past, i.e., for the earlier portions of the audio signal.

The quality condition is implemented such that a switchover to the higher quality encoding algorithm is performed when the quality condition indicates a large quality distance between the first encoding algorithm and the second encoding algorithm. When, for example, it is determined that one encoding algorithm outperforms the other encoding algorithm by, for example, one dB SNR difference, then the quality condition determines a switchover or, stated differently, the actually used encoding algorithm for the actually considered portion of the audio signal irrespective of any transient detection or hysteresis situation.

When, however, the quality condition only indicates a small quality distance between both encoding algorithms such as the quality distance of one or less dB SNR difference, a

switch over to the lower quality encoding algorithm may occur, when the transient detection result indicates that the lower quality encoding algorithm fits to the audio signal characteristic, i.e., whether the audio signal is transient or not. When, however, the transient detection result indicates that the lower quality encoding algorithm does not fit to the audio signal characteristic, then the higher quality encoding algorithm is to be used. In the latter case, once again, the quality condition determines the result, but only when a specific match between the lower quality encoding algorithm and the transient/stationary situation of the audio signal do not fit together.

The hysteresis condition is particularly useful in a combination with the transient condition, i.e., in that the switch to the lower quality encoding algorithm is only performed when less than the last N frames have been encoded with the other algorithm. In preferred embodiments, N is equal to five frames, but other values preferably lower or equal to N frames or signal portions, each comprising a minimum number of samples above e.g. 128 samples, can be used as well.

Fig. 4 illustrates a table of state changes depending on certain situations. The left column indicates the situation where the number of earlier frames is greater than N or smaller than N for either TCX or ACELP.

The last line indicates whether there is a large quality distance for TCX or a large quality distance for ACELP. In these two cases, which are the first two columns, a change is performed where indicated by an "X", while a change is not performed as indicated by "0".

Furthermore, the last two columns indicate the situation when a small quality distance for TCX is determined and when a transient signal is detected or when a small quality distance for an ACELP is determined and the signal portion is detected as being non-transient.

The first two lines of the last two columns both indicate that the quality result is decisive when the number of earlier frames is greater than 10. Hence, when there is a strong indication from the past for one coding algorithm, then the transient detection does not play a role, either.

When, however, the number of earlier frames being encoded in one of the two encoding algorithms is smaller than N, a switchover is performed from TCX to ACELP indicated at field 41 for transient signals. Additionally, as indicated in field 44, a change from ACELP to TCX is performed even when there is a small quality distance in favor of ACELP due to the fact that we have a non-transient signal. When the number of the last LCLP frames is

smaller than N the subsequent frame is also encoded with ACELP and, therefore, no switchover is necessary as indicated at field 42. When, additionally, the number of TCX frames is smaller than N and when there is a small quality distance for ACELP and the signal is non-transient, the current frame is encoded using TCX and, no switchover is necessary as indicated by field 43. Hence, the influence of the hysteresis is clearly visible by comparing fields 42, 43 with the four fields above these two fields.

Hence, the present invention preferably influences the hysteresis for the closed-loop decision by the output of a transient detector. Therefore, there does not exist, as in AMR-WB+, a pure closed-loop decision whether TCX or ACELP is taken. Instead, the closed-loop calculation is influenced by the transient detection result, i.e., every transient signal portion is determined in the audio signal. The decision whether an ACELP frame or TCX frame is calculated, therefore does not only depend on the closed-loop calculations, or, generally, the quality result, but additionally depends on whether a transient is detected or not.

In other words, the hysteresis for determining which encoding algorithm is to be used for the current frame can be expressed as follows:

When the quality result for TCX is slightly smaller than the quality result for ACELP, and when the currently considered signal portions or just the current frame is not transient, then TCX is used instead of ACELP.

When, on the other hand, the quality result for ACELP is slightly smaller than the quality result for TCX, and when the frame is transient, then ACELP is used instead of TCX. Preferably, a flatness measure is calculated as the transient detection result, which is a quantitative number. When the flatness is greater than or equal to a certain value, then the frame is determined to be transient. When, on the other hand, the flatness is smaller than this threshold value, then it is determined that the frame is non-transient. As a threshold, the flatness measure of two is preferred, where the calculation of the flatness is described in Fig. 5 in more detail.

Furthermore, as to the quality result, a quantitative measure is preferred. When an SNR measure or, particularly, a segmental SNR measure is used, then the term "slightly smaller" as used before, may mean one dB smaller. Hence, when the SNRs for TCX and ACELP are more different from each other or stated differently, when the absolute difference between both SNR values is greater than one dB, then the quality condition of Fig. 3 alone determines the encoding algorithm for the current audio signal portion.

The above described decision can be furthermore elaborated, when the transient detection or the hysteresis output or the SNR of TCX or ACELP of the past or earlier frames is included into the if condition. Hence, a hysteresis is built which, for one embodiment, is illustrated in Fig. 3 as condition no. 3. Particularly, Fig. 3 illustrated the alternative when the hysteresis output, i.e., the determination for the past is used for modifying the transient condition.

Alternatively, a further hysteresis condition being based on the earlier TCX or ACELP-SNRs may comprise that a determination for the lower quality encoding algorithm is only performed when a change of the SNR difference with respect to the earlier frame is lower than, for example, a threshold. A further embodiment may comprise the usage of the transient detection result for one or more earlier frames when the transient detection result is a quantitative number. Then, a switchover to the lower quality encoding algorithm may, for example, only be performed when a change of quantitative transient detection result from the earlier frame to the current frame is, again, below a threshold. Other combinations of these figures for further modifying the hysteresis condition 3 of Fig. 3 can prove to be useful in order to obtain a better compromise between the bitrate on the one hand and the audio quality on the other hand.

Furthermore, the hysteresis condition as illustrated in the context of Fig. 3 and as described before can be used instead of or in addition to a further hysteresis which, for example, is based on internal analysis data of the ACELP and TCX encoding algorithms.

Subsequently, reference is made to Fig. 5 for illustrating the preferred determination of the transient detection result on line 14 of Fig. 1.

In step 50, the time-domain audio signal such as a PCM input signal on line 10 is high-pass filtered to obtain a high-pass filtered audio signal. Then, in step 52, the frame of the high-pass filtered signal which can be equal to the portion of the audio signal is sub-divided into a plurality of, for example, eight sub-blocks. Then, in step 54, an energy value for each sub-block is calculated. This energy calculation can comprise a squaring of each sample value in the sub-block and a subsequent addition of the squared samples with or without an averaging. Then, in step 56, pairs of adjacent sub-blocks are formed. The pairs can comprise a first pair consisting of the first and the second sub-block, a second pair consisting of the second and third sub-block, a third pair consisting of the third and fourth sub-block, etc. Additionally, a pair comprising the last sub-block of the earlier frame and the first sub-block of the current frame can be used as well. Alternatively, other ways of

forming pairs can be performed such as, for example, only forming pairs of the first and second sub-block, of the third and fourth sub-block, etc. Then, as also outlined in block 56 of Fig. 5, the higher energy value of each sub-block pair is selected and, as outlined in step 58, divided by the lower energy value of the sub-block pair. Then, as outlined in block 60 of Fig. 5, all results of step 58 for a frame are combined. This combination may consist of an addition of the results of block 58 and an averaging where the result of the addition is divided by the number of pairs such as eight, when eight pairs per sub-block were determined in block 56. The result of block 60 is the flatness measure which is used by the controller 22 in order to determine whether a signal portion is transient or not. When the flatness measure is greater than or equal to 2, a transient signal portion is detected, while, when the flatness measure is lower than 2, it is determined that a signal is non-transient or stationary. However, other thresholds between 1.5 and 3 can be used as well, but it has been shown that the threshold of two provides the best results.

It is to be noted that other transient detectors can be used as well. Transient signals may additionally comprise voiced speech signals. Traditionally, transient signals comprise applause like signals or castagnets or speech plosives comprising signals obtained by speaking characters "p" or "t" or the like. However, vocals such as "a", "e", "i", "o", "u" are not meant to be transient signals in the classical approach, since same are characterized by periodic glottal or pitch pulses. However, since vocals also represent voiced speech signals, vocals are also considered to be transient signals for the present invention. The detection of those signals can be done, in addition or alternative to the procedure in Fig. 5, by speech detectors distinguishing voiced speech from unvoiced speech or by evaluating metadata associated with an audio signal and indicating, to a metadata evaluator, whether the corresponding portion is a transient or non-transient portion.

Subsequently, Fig. 6a is described in order to illustrate the third way of calculating the quality result on line 20 of Fig. 1, i.e., how the processor 18 is preferably configured.

In block 61, a closed-loop procedure is described where, for each of a plurality of possibilities, a portion is encoded and decoded using the first and second coding algorithms. Then, in step 63, a measure such as a segmental SNR is calculated depending on the difference of the encoded and again decoded audio signal and the original signal. This measure is calculated for both encoding algorithms.

Then, an average segmental SNR using the individually segmental SNRs is calculated in step 65, and this calculation is again performed for both encoding algorithms so that, in the end, step 65 results in two different averaged SNR values for the same portion of the audio

signal. The difference between these segmented SNR values for a frame is used as the quantitative quality result on line 20 of Fig. 1.

Fig. 6b illustrates two equations, where the upper equation is used in block 63, and where the lower equation is used in block 65. $\hat{x}_w$ stands for the weighted audio signal, and $\hat{x}_w$ stands for the encoded and again decoded weighted signal.

The averaging performed in block 65 is an averaging over one frame, where each frame consists of a number of subframes N_{SF} , and where four such frames together form a superframe. Hence, a superframe comprises 1024 samples, an individual frame comprises 2056 samples, and each subframe, for which the upper equation in Fig. 6b or step 63 is performed, comprises 64 samples. In the upper equation used in block 63, n is the sample number index and N is the maximum number of samples in the subframe equal to 63 indicating that a subframe has 64 samples.

15

Fig. 7 illustrates a further embodiment of the inventive apparatus for encoding, similar to the Fig. 1 embodiment, and the same reference numerals indicate similar elements. However, Fig. 7 illustrates a more detailed representation of the encoder stage 16, which comprises a pre-processor 16a for performing a weighting and LPC analysis/filtering, and the pre-processor block 16a provides LPC data on line 70 to the output interface 24. Furthermore, the encoder stage 16 of Fig. 1 comprises the first encoding algorithm at 16b and the second encoding algorithm at 16c which are the ACELP encoding algorithm and the TCX encoding algorithm, respectively.

20

Furthermore, the encoder stage 16 may comprise either a switch 16d connected before the blocks 16b, 16c or a switch 16e connected subsequent to the blocks 16b, 16c, where “before” and “subsequent” refer to the signal flow direction which is at least with respect to block 16a to 16e from top to bottom of Fig. 7. Block 16d will not be present in a closed-loop decision. In this case, only switch 16e will be present, since both encoding algorithms 16b, 16c operate on one and the same portion of the audio signal and the result of the selected encoding algorithm will be taken out and forwarded to the output interface 24.

30

If, however, an open-loop decision or any other decision is performed before both encoding algorithms operate on one and the same signal, then switch 16e will not be present, but the switch 16d will be present, and each portion of the audio signal will only be encoded using either one of blocks 16b, 16c.

35

Furthermore, particularly for the closed-loop mode, the outputs of both blocks are connected to the processor and controller block 18, 22 as indicated by lines 71, 72. The switch control takes place via lines 73, 74 from the processor and controller block 18, 22 to the corresponding switches 16d, 16e. Again, depending on the implementation, only one of
5 lines 73, 74 will typically be there.

The encoded audio signal 26 therefore, comprises, among other data, the result of an ACELP or TCX which will typically be redundancy-encoded in addition such as by Huffman-coding or arithmetic coding before being input into the output interface 24.
10 Additionally, the LPC data 70 are provided to the output interface 24 in order to be included in the encoded audio signal. Furthermore, it is preferred to additionally include a coding mode decision into the encoded audio signal indicating to a decoder that the current portion of the audio signal is an ACELP or a TCX portion.

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

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

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

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

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

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

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

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

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

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

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

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

Claims

1. Apparatus for coding a portion of an audio signal to obtain an encoded audio signal for the portion of the audio signal, comprising:
 - 5 a transient detector for detecting whether a transient signal is located in the portion of the audio signal to obtain a transient detection result;
 - an encoder stage for performing a first encoding algorithm on the audio signal, the first encoding algorithm having a first characteristic, and for performing a second encoding algorithm on the audio signal, the second encoding algorithm having a second characteristic being different from the first characteristic;
 - 10 a processor for determining which encoding algorithm results in an encoded audio signal being a better approximation to the portion of the audio signal with respect to the other encoding algorithm to obtain a quality result; and
 - 15 a controller for determining whether the encoded audio signal for the portion of the audio signal is to be generated by either the first encoding algorithm or the second encoding algorithm based on the transient detection result and the quality result, wherein the controller is configured for applying a hysteresis processing so that the second encoding algorithm or the first encoding algorithm is only determined when the lower quality result indicates a lower quality for the second encoding algorithm or the first encoding algorithm, when a number of earlier signal portions having the first encoding algorithm or the second encoding algorithm, respectively, is equal or lower than a predetermined number, and when the transient detection result indicates a predefined state of the two possible states comprising non-transients and transients.
 - 20
 - 25
2. Apparatus in accordance with claim 1, wherein the encoder stage is configured for using a first encoding algorithm which is better suited for the transient signal than the second encoding algorithm.
3. Apparatus of claim 1, wherein the first encoding algorithm is an ACELP coding algorithm, and wherein the second encoding algorithm is a transform coding algorithm.
- 30
- 35

4. Apparatus in accordance with claim 1, wherein the transient detector is configured to perform the following steps:
- high-pass filtering of the audio signal to obtain a high-pass filtered signal block;
- 5 subdividing of the high-pass filtered signal block into a plurality of sub-blocks;
- calculating an energy for each sub-block;
- 10 combining of the energy values for each pair of adjacent sub-blocks to obtain a result for each pair; and
- combining of the results for the pairs to obtain the transient detection result.
- 15 5. Apparatus in accordance with claim 1, wherein the encoder stage further comprises an LPC filtering stage for determining LPC coefficients from the audio signal for filtering the audio signal using an LPC analysis filter determined by the LPC coefficients to determine a residual signal, wherein the first encoding algorithm or the second encoding algorithm is applied to the residual signal, and
- 20 wherein the encoded audio signal further comprises information on the LPC coefficients.
6. Apparatus in accordance with claim 1,
- 25 wherein the encodier stage either comprises a switch connected to the first encoding algorithm and the second encoding algorithm or a switch connected subsequently to the first encoding algorithm and the second encoding algorithm, wherein the switch is controlled by the controller.
- 30 7. Method of coding a portion of an audio signal to obtain an encoded audio signal for the portion of the audio signal, comprising:
- 35 detecting whether a transient signal is located in the portion of the audio signal to obtain a transient detection result;
- performing a first encoding algorithm on the audio signal, the first encoding algorithm having a first characteristic, and performing a second encoding algorithm

on the audio signal, the second encoding algorithm having a second characteristic being different from the first characteristic;

5 determining which encoding algorithm results in an encoded audio signal being a better approximation to the portion of the audio signal with respect to the other encoding algorithm to obtain a quality result; and

10 determining whether the encoded audio signal for the portion of the audio signal is to be generated by either the first encoding algorithm or the second encoding algorithm based on the transient detection result and the quality result, wherein the determining comprises applying a hysteresis processing so that the second encoding
15 algorithm or the first encoding algorithm is only determined when the lower quality result indicates a lower quality for the second encoding algorithm or the first encoding algorithm, when a number of earlier signal portions having the first encoding algorithm or the second encoding algorithm, respectively, is equal or lower than a predetermined number, and when the transient detection result indicates a predefined state of the two possible states comprising non-transients and transients.

20 8. Physical storage medium having stored thereon a machine executable code for performing, when running on a computer, the method of coding a portion of an audio signal in accordance with claim 7.

1/7

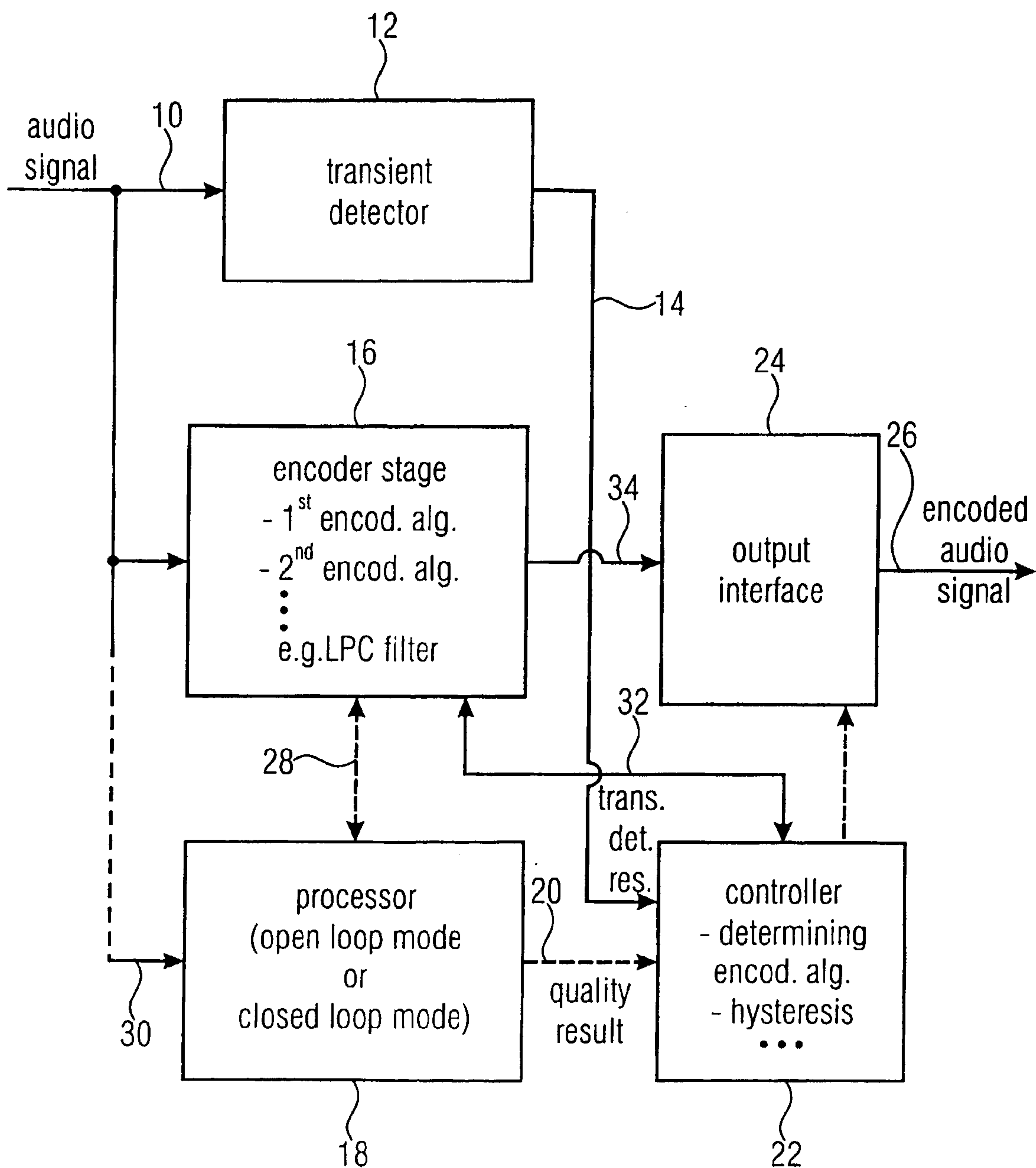


FIGURE 1

encoding algorithm	transient signals	non-transient signals
ACELP	X	
TCX 20		X

FIGURE 2

3/7

ACELP ——— TCX20 switch-over

1. Quality Condition

If quality condition indicates
high quality distance (e.g. more than 1dB SNR)
switch over to higher quality encoding algorithm.

2. Transient Condition

If quality condition indicates
small quality distance (e.g. 1 or less dB SNR)
switch over to lower quality encoding algorithm,
when transient detection result indicates that lower quality
encoding algorithm fits to the audio signal characteristic
otherwise switch to higher quality encoding algorithm.

3. Hysteresis condition

Only switch to lower quality encoding algorithm,
when less than the last N frames have been encoded
with the other algorithm (e.g. N=5).

FIGURE 3

earlier frames	41				42	
> N TCX	0	X	0	X	low quality distance for TCX and transient	low quality distance for ACELP and non-transient
> N ACELP	X	0	X	0		
< N TCX	0	X	X	0	43	44
< N ACELP	X	0	0	X		
	high quality distance for TCX	high quality distance for ACELP	low quality distance for TCX and transient		low quality distance for ACELP and non-transient	

X: switch-over
0: no switch-over

FIGURE 4

FIGURE 4

5/7

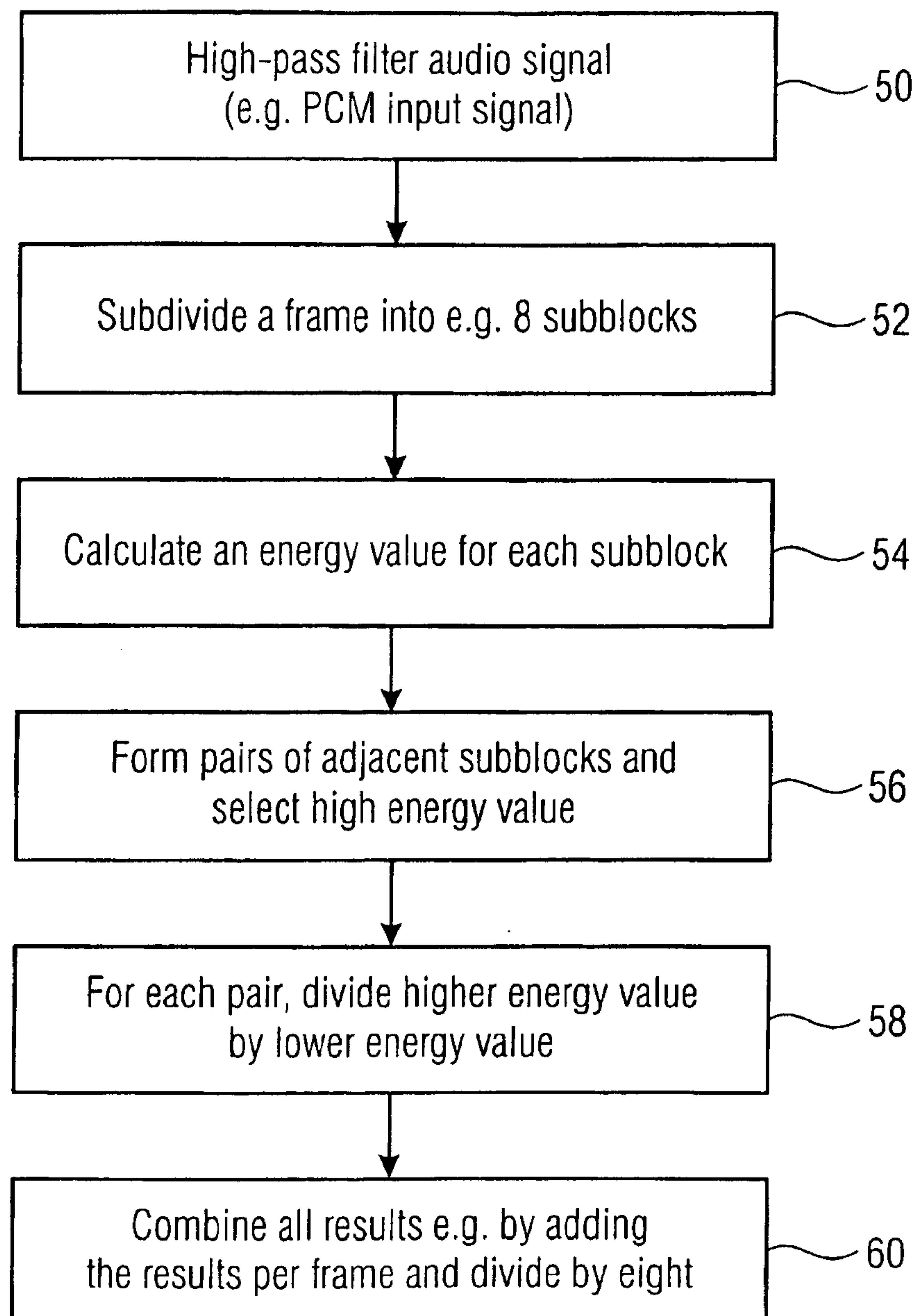


FIGURE 5

6/7

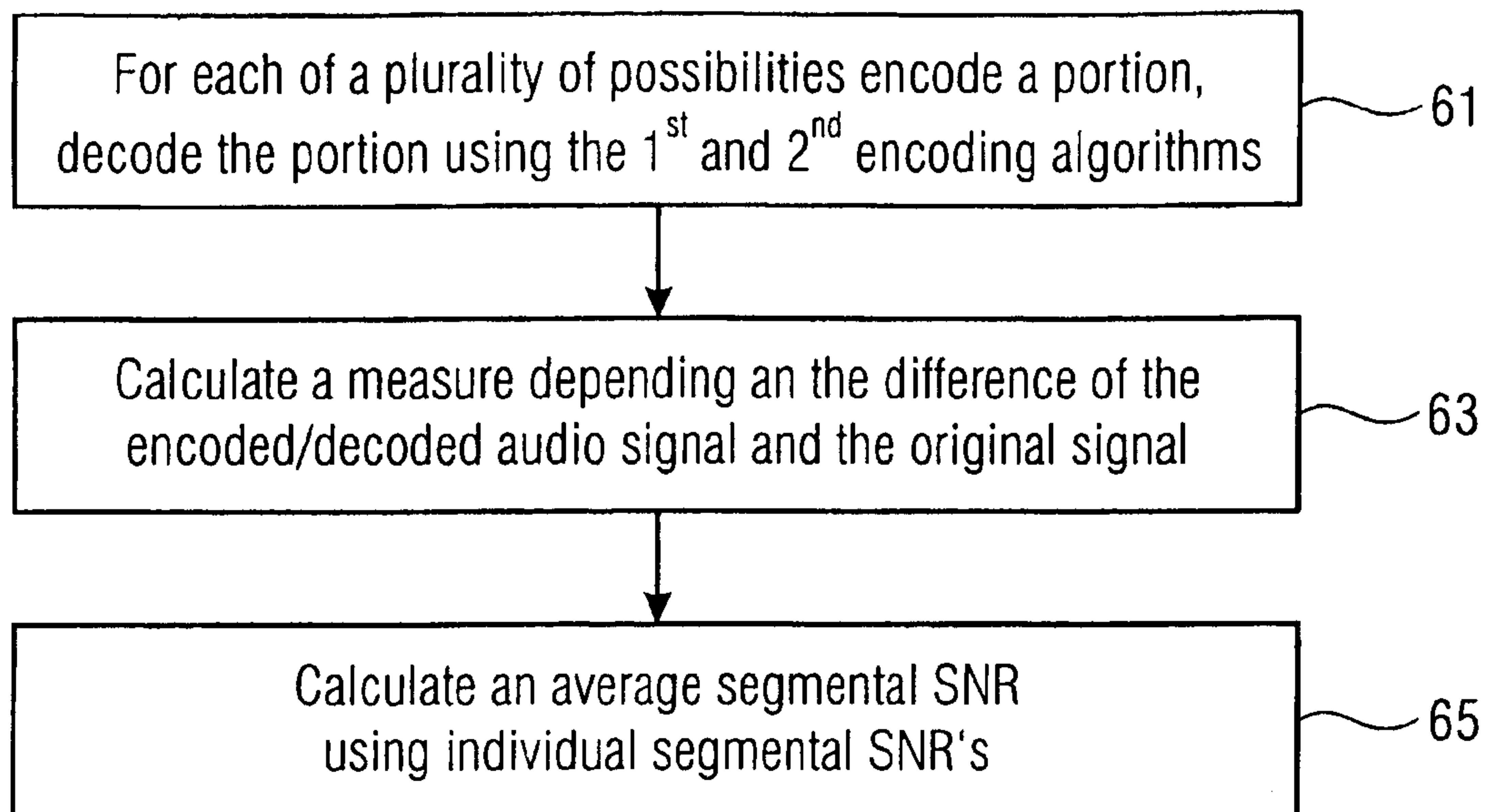


FIGURE 6A

$$\text{segSNR}_i = 20 \log_{10} \left(\frac{\sum_{n=0}^{N-1} x_w^2(n)}{\sum_{n=0}^{N-1} (x_w(n) - \bar{x}_w(n))^2} \right)$$

x_w : wighted audio signal
 $\bar{x}_w$: encoded/decoded weighted signal

$$\overline{\text{segSNR}} = \frac{1}{N_{\text{SF}}} \sum_{i=0}^{N_{\text{SF}}-1} \text{segSNR}_i$$

1 superframe has four frames;
 (1024) (256)
 1 frame has four subframes;
 (256) (64)

N_{SF} = number of
 subframes in a frame

i = sample number index

$N = 64 \hat{=}$ number of
 samples in a subframe

FIGURE 6B

7/7

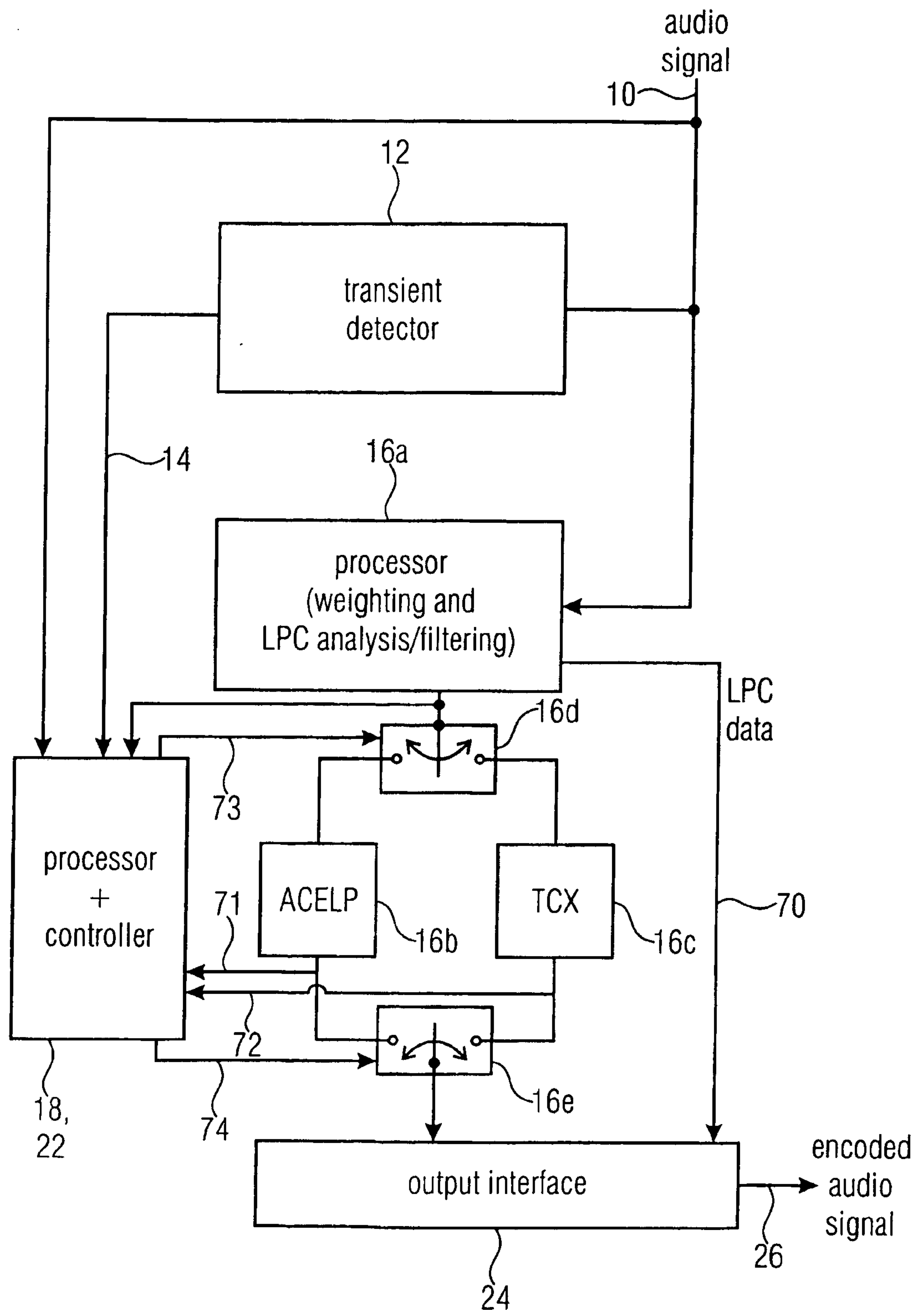


FIGURE 7

