Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2566372 A1 2005/11/24

(21) 2 566 372

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2004/05/17
(87) Date publication PCT/PCT Publication Date: 2005/11/24
(85) Entrée phase nationale/National Entry: 2006/11/09
(86) N° demande PCT/PCT Application No.: IB 2004/001579
(87) N° publication PCT/PCT Publication No.: 2005/112004

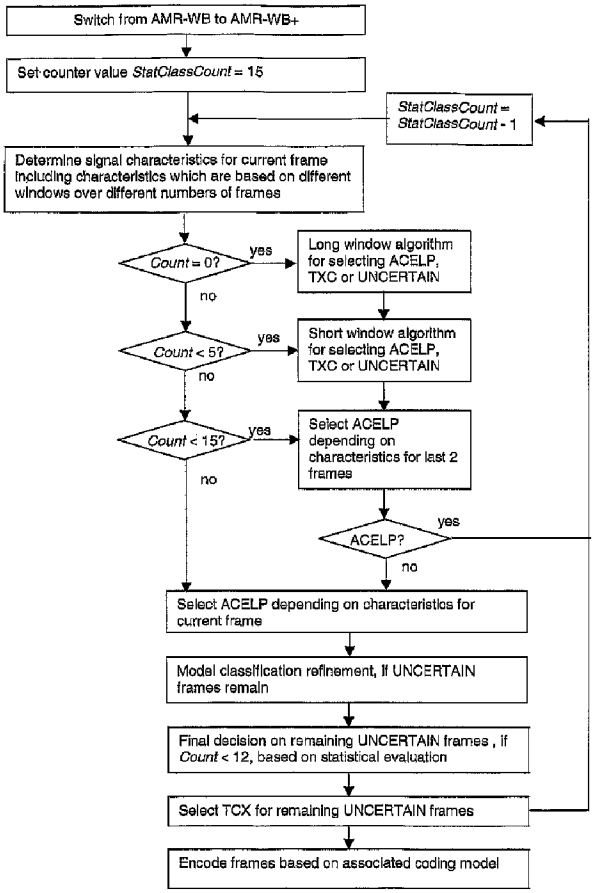
(51) Cl.Int./Int.Cl. *G10L 19/14* (2006.01)

(71) Demandeur/Applicant:
NOKIA CORPORATION, FI

(72) Inventeurs/Inventors:
MAEKINEN, JARI, FI;
LAKANIEMI, ARI, FI;
OJALA, PASI, FI

(74) Agent: SIM & MCBURNEY

(54) Titre : CODAGE AUDIO AVEC DIFFERENTS MODELES DE CODAGE
(54) Title: AUDIO ENCODING WITH DIFFERENT CODING MODELS



(57) Abrégé/Abstract:
The invention relates to a method for supporting an encoding of an audio signal, wherein at least a first and a second coder mode are available for encoding a section of the audio signal. The first coder mode enables a coding based on two different coding

(21) 2 566 372

(13) A1

(57) **Abrégé(suite)/Abstract(continued)**:
models. A selection of a coding model is enabled by a selection rule which is based on signal characteristics which have been determined for a certain analysis window. In order to avoid a misclassification of a section after a switch to the first coder mode, it is proposed that the selection rule is activated only when sufficient sections for the analysis window have been received. The invention relates equally to a module 2,3 in which this method is implemented, to a device 1 and a system comprising such a module 2,3, and to a software program product including a software code for realizing the proposed method.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 November 2005 (24.11.2005)

PCT

(10) International Publication Number
WO 2005/112004 A1

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

(74) Agent: COHAUSZ & FLORACK (24); Bleichstraße 14, 40211 Düsseldorf (DE).

(21) International Application Number:
PCT/IB2004/001579

(22) International Filing Date: 17 May 2004 (17.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **NOKIA CORPORATION** [FI/FI]; Keilalahdentie 4, FIN-02150 Espoo (FI).

(72) Inventors; and

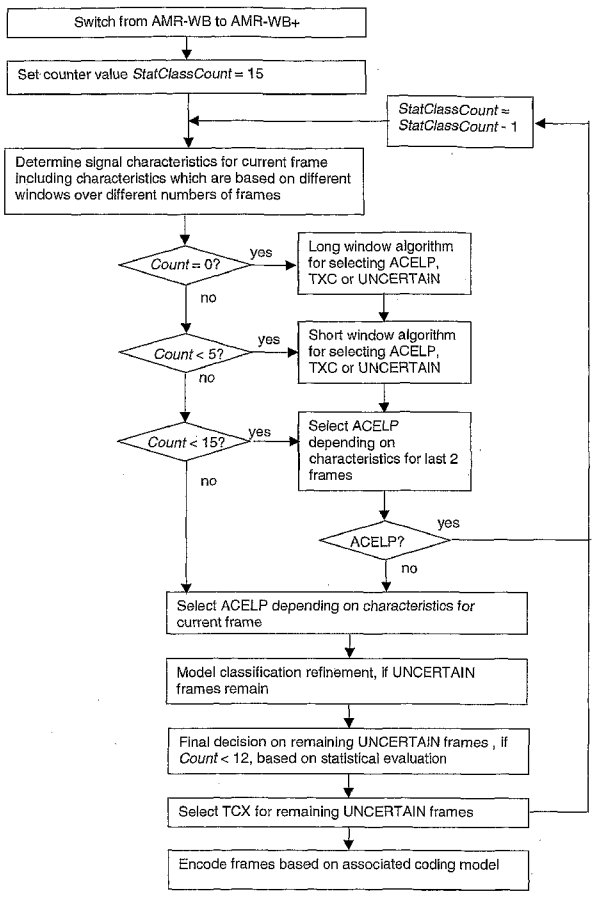
(75) Inventors/Applicants *(for US only)*: **MÄKINEN, Jari** [FI/FI]; Etuniitynkatu 4 as. 2, FIN-33580 Tampere (FI). **LAKANIEMI, Ari** [FI/FI]; Kalliolanrinne 6 C 66, FIN-00510 Helsinki (FI). **OJALA, Pasi** [FI/FI]; Gresantie 9 A 2, FIN-02700 Kauniainen (FI).

(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, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, IT, LU, MC, NL, PL, PT, RO, SE, SI,

[Continued on next page]

(54) Title: AUDIO ENCODING WITH DIFFERENT CODING MODELS



(57) Abstract: The invention relates to a method for supporting an encoding of an audio signal, wherein at least a first and a second coder mode are available for encoding a section of the audio signal. The first coder mode enables a coding based on two different coding models. A selection of a coding model is enabled by a selection rule which is based on signal characteristics which have been determined for a certain analysis window. In order to avoid a misclassification of a section after a switch to the first coder mode, it is proposed that the selection rule is activated only when sufficient sections for the analysis window have been received. The invention relates equally to a module 2,3 in which this method is implemented, to a device 1 and a system comprising such a module 2,3, and to a software program product including a software code for realizing the proposed method.

WO 2005/112004 A1

WO 2005/112004 A1

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.

Audio encoding with different coding models

FIELD OF THE INVENTION

The invention relates to a method for supporting an encoding of an audio signal, wherein at least a first coder mode and a second coder mode are available for encoding a specific section of the audio signal. At least the first coder mode enables a coding of a specificsection of the audio signal based on at least two different coding models. In the first coder mode a selection of a respective coding model for encoding a specific section of an audio signal is enabled by at least one selection rule which is based on an analysis of signal characteristics in an analysis window which covers at least one section of the audio signal preceding the specific section. The invention relates equally to a corresponding module, to a corresponding electronic device, to a corresponding system and to a corresponding software program product.

BACKGROUND OF THE INVENTION

It is known to encode audio signals for enabling an efficient transmission and/or storage of audio signals.

An audio signal can be a speech signal or another type of audio signal, like music, and for different types of audio signals different coding models might be appropriate.

A widely used technique for coding speech signals is the Algebraic Code-Excited Linear Prediction (ACELP) coding. ACELP models the human speech production system, and it is very well suited for coding the periodicity of a speech signal. As a result, a high speech quality can be achieved with very low bit rates. Adaptive Multi-Rate Wideband (AMR-WB), for example, is a speech codec which is based on the ACELP technology. AMR-WB has been described for instance in the technical specification 3GPP TS 26.190: "Speech Codec speech processing functions; AMR Wideband speech codec; Transcoding functions", V5.1.0 (2001-12). Speech codecs which are based on the human speech production system, however, perform usually rather badly for other types of audio signals, like music.

A widely used technique for coding other audio signals than speech is transform coding (TCX). The superiority of transform coding for audio signal is based on perceptual masking and frequency domain coding. The quality of the resulting audio signal can be further improved by selecting a suitable coding frame length for the transform coding. But while transform coding techniques result in a high quality for audio signals other than speech, their performance is not good for periodic speech signals. Therefore, the quality of transform coded speech is usually rather low, especially with long TCX frame lengths.

The extended AMR-WB (AMR-WB+) codec encodes a stereo audio signal as a high bitrate mono signal and provides some side information for a stereo extension. The AMR-WB+ codec utilizes both, ACELP coding and TCX models to encode the core mono signal in a frequency band of 0 Hz

to 6400 Hz. For the TCX model, a coding frame length of 20 ms, 40 ms or 80 ms is utilized.

Since an ACELP model can degrade the audio quality and transform coding performs usually poorly for speech, especially when long coding frames are employed, the respectively best coding model has to be selected depending on the properties of the signal which is to be coded. The selection of the coding model that is actually to be employed can be carried out in various ways.

In systems requiring low complex techniques, like mobile multimedia services (MMS), usually music/speech classification algorithms are exploited for selecting the optimal coding model. These algorithms classify the entire source signal either as music or as speech based on an analysis of the energy and the frequency properties of the audio signal.

If an audio signal consists only of speech or only of music, it will be satisfactory to use the same coding model for the entire signal based on such a music/speech classification. In many other cases, however, the audio signal that is to be encoded is a mixed type of audio signal. For example, speech may be present at the same time as music and/or be temporally alternating with music in the audio signal.

In these cases, a classification of entire source signals into music or speech category is a too limited approach. The overall audio quality can then only be maximized by temporally switching between the coding models when coding the audio signal. That is, the ACELP model is partly used as well for coding a source signal classified

as an audio signal other than speech, while the TCX model is partly used as well for a source signal classified as a speech signal.

The extended AMR-WB (AMR-WB+) codec is designed as well for coding such mixed types of audio signals with mixed coding models on a frame-by-frame basis.

The selection of coding models in AMR-WB+ can be carried out in several ways.

In the most complex approach, the signal is first encoded with all possible combinations of ACELP and TCX models. Next, the signal is synthesized again for each combination. The best excitation is then selected based on the quality of the synthesized speech signals. The quality of the synthesized speech resulting with a specific combination can be measured for example by determining its signal-to-noise ratio (SNR). This analysis-by-synthesis type of approach will provide good results. In some applications, however, it is not practicable, because of its very high complexity. Such applications include, for example, mobile applications. The complexity results largely from the ACELP coding, which is the most complex part of an encoder.

In systems like MMS, for example, the full closed-loop analysis-by-synthesis approach is far too complex to perform. In an MMS encoder, therefore, a low complex open-loop method is employed for determining whether an ACELP coding model or a TCX model is selected for encoding a particular frame.

AMR-WB+ offers two different low-complex open-loop approaches for selecting the respective coding model for each frame. Both open-loop approaches evaluate source signal characteristics and encoding parameters for selecting a respective coding model.

In the first open-loop approach, an audio signal is first split up within each frame into several frequency bands, and the relation between the energy in the lower frequency bands and the energy in the higher frequency bands is analyzed, as well as the energy level variations in those bands. The audio content in each frame of the audio signal is then classified as a music-like content or a speech-like content based on both of the performed measurements or on different combinations of these measurements using different analysis windows and decision threshold values.

In the second open-loop approach, which is also referred to as model classification refinement, the coding model selection is based on an evaluation of the periodicity and the stationary properties of the audio content in a respective frame of the audio signal. Periodicity and stationary properties are evaluated more specifically by determining correlation, Long Term Prediction (LTP) parameters and spectral distance measurements.

The AMR-WB+ codec allows in addition to switch during the coding of an audio stream between AMR-WB modes, which employ exclusively an ACELP coding model, and extension modes, which employ either an ACELP coding model or a TCX model, provided that the sampling frequency does not change. The sampling frequency can be for example 16 kHz.

The extension modes output a higher bit rate than the AMR-WB modes. A switch from an extension mode to an AMR-WB mode can thus be of advantage when transmission conditions in the network connecting the encoding end and the decoding end require a changing from a higher bit-rate mode to a lower bit-rate mode to reduce congestion in the network. A change from a higher bit-rate mode to a lower bit-rate mode might also be required for incorporating new low-end receivers in a Mobile Broadcast/Multicast Service (MBMS).

A switch from an AMR-WB mode to an extension mode, on the other hand, can be of advantage when a change in the transmission conditions in the network allows a change from a lower bit-rate mode to a higher bit-rate mode. Using a higher bit-rate mode enables a better audio quality.

Since the core codec use the same sampling rate of 6.4kHz for the AMR-WB modes and the AMR-WB+ extension modes and employs at least partially similar coding techniques, a change from an extension mode to an AMR-WB mode, or vice versa, at this frequency band can be handled smoothly. As the core-band coding process is slightly different for an AMR-WB mode and an extension mode, it has to be taken care, however, that all required state variables and buffers are stored and copied from one algorithm to the other when switching between the modes.

Further, it has to be taken into account that a coding model selection is only required in the extension modes. In the enabled open-loop classification approaches, relatively long analysis windows and data buffers are exploited. The encoding model selection exploits

statistical analysis with analysis windows having a length of up to 320 ms, which corresponds to 16 audio signal frames of 20 ms. Since a corresponding information does not have to be buffered in the AMR-WB mode, it cannot simply be copied to the extended mode algorithms. After switching from AMR-WB to AMR-WB+, the data buffers of classification algorithms, for instance those used for a statistical analysis, have thus no valid information or they are reset. During the first 320 ms after a switch, the coding model selection algorithm may thus not be fully adapted or updated for the current audio signal. A selection, which is based on non-valid buffer data results in a distorted coding model decision. For example, an ACELP coding model may be weighted heavily in the selection, even though the audio signal requires a coding based on a TCX model in order to maintain the audio quality.

Thus, the encoding model selection is not optimal, since the low complex coding model selection performs badly after a switch from an AMR-WB mode to an extension mode.

SUMMARY OF THE INVENTION

It is an object of the invention to improve the selection of a coding model after a switching from a first coding mode to a second coding mode.

A method for supporting an encoding of an audio signal is proposed, wherein at least a first coder mode and a second coder mode are available for encoding a specific section of the audio signal. Further, at least the first coder mode enables a coding of a specific section of the audio signal based on at least two different coding

models. In the first coder mode a selection of a respective coding model for encoding a specific section of an audio signal is enabled by at least one selection rule which is based on signal characteristics which have been determined at least partly from an analysis window which covers at least one section of the audio signal preceding the specific section. It is proposed that the method comprising after a switch from the second coder mode to the first coder mode activating the at least one selection rule in response to having received at least as many sections of the audio signal as are covered by the analysis window.

The first coder mode and the second coder mode can be for example, though not exclusively, an extension mode and an AMR-WB mode of an AMR-WB+ codec, respectively. The coding models available for the first coder mode can then be for example an ACELP coding model and a TCX model.

Moreover, a module for supporting an encoding of an audio signal is proposed. The module comprises a first coder mode portion adapted to encode a specific section of an audio signal in a first coder mode and a second coder mode portion adapted to encode a respective section of an audio signal in a second coder mode. The module further comprises switching means for switching between the first coder mode portion and the second coder mode portion. The coder mode portion includes an encoding portion which is adapted to encode a respective section of the audio signal based on at least two different coding models. The first coder mode portion further comprises a selection portion adapted to apply at least one selection rule for selecting a respective coding model, which is to be used by the encoding portion for encoding a specific section

of an audio signal. The at least one selection rule is based on signal characteristics which have been determined at least partly from an analysis window covering at least one section of an audio signal preceding the specific section. The selection portion is adapted to activate the at least one selection rule after a switch by the switching means from the second coder mode portion to the first coder mode portion in response to having received at least as many sections of the audio signal as are covered by the analysis window.

This module can be for instance an encoder or a part of an encoder.

Moreover, an electronic device is proposed, which comprises such a module.

Moreover, an audio coding system is proposed which comprises such a module and in addition a decoder for decoding audio signals which have been encoded by such a module.

Finally, a software program product is proposed, in which a software code for supporting an encoding of an audio signal is stored. At least a first coder mode and a second coder mode are available for encoding a respective section of the audio signal. At least the first coder mode enables a coding of a respective section of the audio signal based on at least two different coding models. In the first coder mode a selection of a respective coding model for encoding a specific section of an audio signal is enabled by at least one selection rule which is based on signal characteristics which have been determined from an analysis window which covers at

least one section of the audio signal preceding the specific section. When running in a processing component of an encoder, the software code activates the at least one selection rule after a switch from the second coder mode to the first coder mode in response to having received at least as many sections of the audio signal as are covered by the analysis window.

The invention proceeds from the consideration that problems with invalid buffer contents which are used as basis for a selection of a coding model can be avoided, if such a selection is only activated after the buffer contents have been updated at least to an extent required by the respective type of selection. It is therefore proposed that when a selection rule uses signal characteristics which have been determined using an analysis window over a plurality of sections of the audio signal, the selection rule is only applied when all sections required by the analysis window have been received. It is to be understood that the activation may be part of the selection rule itself.

It is an advantage of the invention that it enables an improved selection of the coding model after a switch of the coder mode. It allows more specifically to prevent a misclassification of sections of an audio signal, and thus to prevent the selection of an inappropriate coding model.

For the time after a switching in which some selection rules have not been activated, advantageously an additional selection rule is provided which does not use information on sections of the audio signal preceding the current section. This further rule can be applied

immediately after a switching and at least as long until other selection rules have been activated.

The at least one selection rule which is based on signal characteristics which have been determined in an analysis window may comprise a single selection rule or a plurality of selection rules. In the latter case, the associated analysis windows may have different lengths. As a result, the plurality of selection rules may be activated one after the other.

The section of an audio signal can be in particular a frame of an audio signal, for instance an audio signal frame of 20 ms.

The signal characteristics which are evaluated by the at least one selection rule may be based entirely or only partly on an analysis window. It is to be understood that also the signal characteristics employed by a single selection rule may be based on different analysis windows.

BRIEF DESCRIPTION OF THE FIGURES

Other objects and features of the present invention will become apparent from the following detailed description considered in conjunction with the accompanying drawings.

Fig. 1 is a schematic diagram of an audio coding system according to an embodiment of the invention; and

Fig. 2 is a flow chart illustrating an embodiment of the method according to the invention implemented in the system of Figure 1.

DETAILED DESCRIPTION OF THE INVENTION

Figure 1 is a schematic diagram of an audio coding system according to an embodiment of the invention, which allows a soft activation of selection algorithms used for selecting an optimal coding model.

The system comprises a first device 1 including an AMR-WB+ encoder 2 and a second device 21 including an AMR-WB+ decoder 22. The first device 1 can be for instance an MMS server, while the second device 21 can be for instance a mobile phone or some other mobile device.

The AMR-WB+ encoder 2 comprises an AMR-WB encoding portion 4 which is adapted to perform a pure ACELP coding, and an extension encoding portion 5, which is adapted to perform a encoding based either on an ACELP coding model or on a TCX model. The extension encoding portion 5 thus constitutes the first coder mode portion and the AMR-WB encoding portion 4 the second coder mode portion of the invention.

The AMR-WB+ encoder 2 further comprises a switch 6 for forwarding audio signal frames either to the AMR-WB encoding portion 4 or to the extension encoding portion 5.

The extension encoding portion 5 comprises a signal characteristics determination portion 11 and a counter 12. The terminal of the switch 6 which is associated to the extension encoding portion 5 is linked to an input of both portions 11, 12. The output of the signal characteristics determination portion 11 and the output of the counter 12 are linked within the extension

encoding portion 5 via a first selection portion 13, a second selection portion 14, a third selection portion 15, a verification portion 16, a refinement portion 17 and a final selection portion 18 to an ACELP/TCX encoding portion 19.

It is to be understood that the presented portions 11 to 19 are designed for encoding a mono audio signal, which may have been generated from a stereo audio signal. Additional stereo information may be generated in additional stereo extension portions not shown. It is moreover to be noted that the encoder 2 comprises further portions not shown. It is also to be understood that the presented portions 12 to 19 do not have to be separate portions, but can equally be interweaved among each others or with other portions.

The AMR-WB encoding portion 4, the extension encoding portion 5 and the switch 6 can be realized in particular by a software SW run in a processing component 3 of the encoder 2, which is indicated by dashed lines.

The processing in the extension encoding portion 5 will now be described in more detail with reference to the flow chart of Figure 2.

The encoder 2 receives an audio signal, which has been provided to the first device 1. At first, the switch 6 provides the audio signal to the AMR-WB encoding portion 4 for achieving a low output bit-rate, for example because there is not sufficient capacity in the network connecting the first device 1 and the second device 21. Later, however, the conditions in the network change and allow a higher bit-rate. The audio signal is therefore

now forwarded by the switch 6 to the extension encoding portion 5.

In case of such a switch, a value *StatClassCount* of the counter 12 is reset to 15 when the first audio signal frame is received. In the following the counter 12 decrements its value *StatClassCount* by one, each time a further audio signal frame is input to the extension encoding portion 5.

Moreover, the signal characteristics determination portion 11 determines for each input audio signal frame various energy related signal characteristics by means of AMR-WB Voice Activity Detector (VAD) filter banks.

For each input audio signal frame of 20 ms, the filter banks produce the signal energy $E(n)$ in each of twelve non-uniform frequency bands covering a frequency range from 0 Hz to 6400 Hz. The energy level $E(n)$ of each frequency band n is then divided by the width of this frequency band in Hz, in order to produce a normalized energy level $E_N(n)$ for each frequency band.

Next, the respective standard deviation of the normalized energy levels $E_N(n)$ is calculated for each of the twelve frequency bands using on the one hand a short window $std_{short}(n)$ and on the other hand a long window $std_{long}(n)$. The short window has a length of four audio signal frames, and the long window has a length of sixteen audio signal frames. That is, for each frequency band, the energy level from the current frame and the energy level from the preceding 4 and 16 frames, respectively, are used to derive the two standard deviation values. The normalized energy levels of the preceding frames are

retrieved from buffers, in which also the normalized energy levels of the current audio signal frame are stored for further use.

The standard deviations are only determined, however, if a voice activity indicator VAD indicates active speech for the current frame. This will make the algorithm react faster especially after long speech pauses.

Now, the determined standard deviations are averaged over the twelve frequency bands for both, long and short window, to create two average standard deviation values $stda_{short}$ and $stda_{long}$ as a first and a second signal characteristic for the current audio signal frame.

For the current audio signal frame, moreover a relation between the energy in the lower frequency bands and the energy in the higher frequency bands is calculated. To this end, the signal characteristics determination portion 11 sums the energies $E(n)$ of the lower frequency bands n = 1 to 7 to obtain an energy level *LevL*. The energy level *LevL* is normalized by dividing it by the total width of these lower frequency bands in Hz. Moreover, the signal characteristics determination portion 11 sums the energies $E(n)$ of the higher frequency bands n = 8 to 11 to obtain an energy level *LevH*. The energy level *LevH* is equally normalized by dividing it by the total width of the higher frequency bands in Hz. The lowest frequency band 0 is not used in these calculations, because it usually contains so much energy that it will distort the calculations and make the contributions from the other frequency bands too small. Next, the signal characteristics determination portion 11 defines the relation $LPH = LevL / LevH$. In addition, a

moving average *LPHa* is calculated using the *LPH* values which have been determined for the current audio signal frame and for the three previous audio signal frames.

Now, a final value *LPHaF* of the energy relation is calculated for the current frame by summing the current *LPHa* value and the previous seven *LPHa* values. In this summing, the latest values of *LPHa* are weighted slightly higher than the older values of *LPHa*. The previous seven values of *LPHa* are equally retrieved from buffers, in which also the value of *LPHa* for the current frame is stored for further use. The value *LPHaF* constitutes the third signal characteristic.

The signal characteristics determination portion 11 calculates in addition an energy average level of the filter banks *AVL* for the current audio signal frame. For calculating the value *AVL*, an estimated level of the background noise is subtracted from the energy *E(n)* in each of the twelve frequency bands. The results are then multiplied with the highest frequency in Hz of the corresponding frequency band and summed. The multiplication allows balancing the influence of the high frequency bands, which contain relatively less energy than the lower frequency bands. The value *AVL* constitutes a fourth third signal characteristic

Finally, the signal characteristics determination portion 11 calculates for the current frame the total energy $TotE_0$ from all filter banks, reduced by an estimate of the background noise for each filter bank. The total energy $TotE_0$ is also stored in a buffer. The value $TotE_0$ constitutes a fifth signal characteristic

The determined signal characteristics and the counter value *StatClassCount* are now provided to the first selection portion 13, which applies an algorithm according to the following pseudo-code for selecting the best coding model for the current frame:

```
if (StatClassCount == 0)
  SET TCX_MODE
if (stda_long < 0.4)
    SET TCX_MODE
else if (LPHaF > 280)
    SET TCX_MODE
else if ( stda_long >= 0.4)
    if ((5+(1/( stda_long -0.4))) > LPHaF)
      SET TCX_MODE
    else if ((-90* stda_long +120) < LPHaF)
      SET ACELP_MODE
    else
      SET UNCERTAIN_MODE
else
  headMode = UNCERTAIN_MODE
```

It can be seen that this algorithm exploits a signal characteristic $stda_{long}$, which is based on information on sixteen preceding audio signal frames. Therefore, it is checked first whether at least seventeen frames have already been received after the switch from AMR-WB. This is the case as soon as the counter 12 has a value *StatClassCount* of zero. Otherwise, an uncertain mode is associated immediately to the current frame. This ensures that the result is not falsified by invalid buffer contents resulting in incorrect values for signal characteristics $stda_{long}$ and *LPHaF*.

Information on the signal characteristics and the coding model selection performed so far is now forwarded by the first selection portion 13 to the second selection portion 14, which applies an algorithm according to the following pseudo-code for selecting the best coding model for the current frame:

```
if (ACELP_MODE or UNCERTAIN_MODE) and (AVL > 2000)
  SET TCX_MODE
if (StatClassCount < 5)
if (UNCERTAIN_MODE)
    if (stda_short < 0.2)
      SET TCX_MODE
    else if (stda_short >= 0.2)
      if ((2.5+(1/( stda_short -0.2))) > LPHaF)
        SET TCX_MODE
      else if ((-90* stda_short+140) < LPHaF)
        SET ACELP_MODE
      else
        SET UNCERTAIN_MODE
```

It can be seen that the second part of this algorithm exploits a signal characteristic $stda_{short}$, which is based on information on four preceding audio signal frames, and moreover a signal characteristic LPHaF, which is based on information on ten preceding audio signal frames. For this part of the algorithm it is therefore checked first whether at least eleven frames have already been received after the switch from AMR-WB. This is the case as soon as the counter has a value *StatClassCount* of '4'. This ensures that the result is not falsified by invalid buffer contents resulting in incorrect values for signal characteristics LPhaF and $stda_{short}$. On the whole, this algorithm allows a selection of a coding model already

for the eleventh to sixteenth frame, and in addition even for the first ten frames in case the average energy level *AVL* exceeds a predetermined value. This part of the algorithm is not indicated in Figure 2. The algorithm is equally applied for frames succeeding the sixteenth frame for refining the first selection by the first selection portion 13.

Information on the signal characteristics and the coding model selection performed so far is then forwarded by the second selection portion 14 to the third selection portion 15, which applies an algorithm according to the following pseudo-code for selecting the best coding model for the current frame, if the mode for this frame is still uncertain:

```
if (UNCERTAIN_MODE)
  if (StatClassCount < 15)
    if ((TotE0 / TotE-1)>25)
      SET ACELP_MODE
```

It can be seen that this pseudo-code exploits the relation between the total energy $TotE_0$ in the current audio signal frame and the total energy $TotE_{-1}$ in the preceding audio signal frame. It is therefore checked first, whether at least two frames have already been received after the switch from AMR-WB. This is the case as soon as the counter has a value *StatClassCount* of '14'.

It has to be noted that the employed counter threshold values are only examples and might be selected in many different ways. In the algorithm implemented in the second selection portion 14, for instance, the signal

characteristic LPH could be evaluated instead of the signal characteristic LPHaF. In this case, it would be sufficient to check whether at least five frames have already been received, corresponding to *StatClassCount* < 12.

Information on the signal characteristics and the coding model selection performed so far is then forwarded by the third selection portion 15 to the verification portion 16, which applies an algorithm according to the following pseudo-code:

```
if (TCX_MODE || UNCERTAIN_MODE))
  if (AVL > 2000 and TotE0 < 60)
    SET ACELP_MODE
```

This algorithm allows selecting possibly the best coding model for the current frame, if the mode for this frame is still uncertain, and to verifying whether an already selected TCX mode is appropriate.

Also after the processing in the verification portion 16, the mode associated to the current audio signal frame may still be uncertain.

In a fast approach, now simply a predetermined coding model, that is either an ACELP coding model or a TCX coding model, is selected for the remaining UNCERTAIN mode frames.

In a more sophisticated approach, illustrated as well in Figure 2, some further analysis is performed first.

To this end, information on the coding model selection performed so far is now forwarded by the verification portion 16 to the refinement portion 17. The refinement portion 17 applies a model classification refinement. As mentioned above, this is a coding model selection, which is based on the periodicity and the stationary properties of the audio signal. The periodicity is observed by using LTP parameters. The stationary properties are analyzed by using a normalized correlation and spectral distance measurements.

The analysis by portions 13, 14, 15, 16 and 17 determine based on audio signal characteristics whether the content of a respective frame can be assumed to be speech or other audio content, like music, and selected a corresponding coding model if such a classification is possible. Portions 13, 14, 15, 16 realize a first open loop approach evaluating energy related characteristics, while portion 17 realizes a second open loop approach evaluating periodicity and the stationary properties of the audio signal.

In case two different open loop approaches have been applied in vain to select a TCX model or an ACELP coding model, the optimal encoding model will be difficult to select in some cases by further existing open loop algorithms. In the present embodiment, therefore a simple counting-based classification is employed for the remaining unclear mode selections.

The final selection portion 18 selects a specific coding model for remaining UNCERTAIN mode frames based on a statistical evaluation of the coding models associated to the respective neighboring frames, if a voice activity

indicator VADflag is set for the respective UNCERTAIN mode frame.

For the statistical evaluation, a current superframe, to which an UNCERTAIN mode frame belongs, and a previous superframe preceding this current superframe are considered. A superframe has a length of 80 ms and comprises four consecutive audio frames of 20 ms each. The final selection portion 18 counts by means of counters the number of frames in the current superframe and in the previous superframe for which the ACELP coding model has been selected by one of the preceding selection portions 12 to 17. Moreover, the final selection portion 18 counts the number of frames in the previous superframe for which a TCX model with a coding frame length of 40 ms or 80 ms has been selected by one of the preceding selection portions 12 to 17, for which moreover the voice activity indicator is set, and for which in addition the total energy exceeds a predetermined threshold value. The total energy can be calculated by dividing the audio signal into different frequency bands, by determining the signal level separately for all frequency bands, and by summing the resulting levels. The predetermined threshold value for the total energy in a frame may be set for instance to 60.

The assignment of coding models has to be completed for an entire current superframe, before the current superframe n can be encoded. The counting of frames to which an ACELP coding model has been assigned is thus not limited to frames preceding an UNCERTAIN mode frame. Unless the UNCERTAIN mode frame is the last frame in the current superframe, also the selected encoding models of upcoming frames are take into account.

The counting of frames can be summarized for instance by the following pseudo-code:

```
if ((prevMode(i) == TCX80 or prevMode(i) == TCX40) and
vadFlag_old(i)== 1 and TotE_i > 60)
    TCXCount = TCXCount + 1
if (prevMode(i) == ACELP_MODE)
    ACELPCount = ACELPCount + 1
if (j != i)
    if (Mode(i) == ACELP_MODE)
      ACELPCount = ACELPCount + 1
```

In this pseudo-code, i indicates the number of a frame in a respective superframe, and has the values 1, 2, 3, 4, while j indicates the number of the current frame in the current superframe. *prevMode(i)* is the mode of the i:th frame of 20ms in the previous superframe and *Mode(i)* is the mode of the i:th frame of 20 ms in the current superframe. TCX80 represents a selected TCX model using a coding frame of 80 ms and TCX40 represents a selected TCX model using a coding frame of 40 ms. $vadFlag_{old}(i)$ represents the voice activity indicator VAD for the i:th frame in the previous superframe. $TotE_i$ is the total energy in the i:th frame. The counter value *TCXCount* represents the number of selected long TCX frames in the previous superframe, and the counter value *ACELPCount* represents the number of ACELP frames in the previous and the current superframe.

A statistical evaluation is then performed as follows:

If the counted number of long TCX mode frames, with a coding frame length of 40 ms or 80 ms, in the previous

superframe is larger than 3, a TCX model is equally selected for the UNCERTAIN mode frame.

Otherwise, if the counted number of ACELP mode frames in the current and the previous superframe is larger than 1, an ACELP model is selected for the UNCERTAIN mode frame.

In all other cases, a TCX model is selected for the UNCERTAIN mode frame.

The selection of the coding model Mode(j) for the j:th frame can be summarized for instance by the following pseudo-code:

```
if (TCXCount > 3)
  Mode(j) = TCX_MODE;
else if (ACELPCount > 1)
  Mode(j) = ACELP_MODE
else
  Mode(j) = TCX_MODE
```

The counting-based approach is only performed, if the counter value *StatClassCount* is smaller than 12. This means, that after switching from AMR-WB to an extension mode, the counting-based classification approach is not performed in the first four frames, which is for the first 4*20 ms.

If the counter value *StatClassCount* is equal to or larger than 12 and the encoding model is still classified as UNCERTAIN mode, the TCX model is selected.

If the voice activity indicator VADflag is not set, the flag thereby indicating a silent period, the selected

mode is TCX by default and none of the mode selection algorithms has to be performed.

The portions 13, 14 and 15 thus constitute the at least one selection portion of the invention, while the portions 16, 17 and 18, and partly portion 14, constitute the at least one further selection portion of the invention.

The ACELP/TCX encoding portion 19 now encodes all frames of the audio signal based on the respectively selected coding model. The TCX model is based by way of example on an fast Fourier transform (FFT) using the selected coding frame length, and the ACELP coding model uses by way of example an LTP and fixed codebook parameters for a linear prediction coefficients (LPC) excitation.

The encoding portion 19 then provides the encoded frames for a transmission to the second device 21. In the second device 21, the decoder 22 decodes all received frames with the ACELP coding model or with the TCX coding model using an AMR-WB mode or an extension mode, as required. The decoded frames are provided for example for presentation to a user of the second device 21.

Summarized, the presented embodiment enables a soft activation of selection algorithms, in which the provided selection algorithms are activated in the order in which analysis buffers that are related to the selection rules are fully updated. While one or more selection algorithms are disabled, the selection is performed based on other selection algorithms, which do not rely on this buffer content.

It is to be noted that the described embodiment constitutes only one of a variety of possible embodiments of the invention.

Claims

1. A method for supporting an encoding of an audio signal, wherein at least a first coder mode and a second coder mode are available for encoding a specific section of said audio signal, wherein at least said first coder mode enables a coding of a specific section of said audio signal based on at least two different coding models, and wherein in said first coder mode a selection of a respective coding model for encoding said specific section of an audio signal is enabled by at least one selection rule which is based on signal characteristics, which signal characteristics have at least partly been determined from an analysis window, which analysis window covers at least one section of said audio signal preceding said specific section, said method comprising after a switch from said second coder mode to said first coder mode activating said at least one selection rule in response to having received at least as many sections of said audio signal as are covered by said analysis window.

2. A method according to claim 1, wherein in said first coder mode a selection of a respective coding model for encoding a specific section of an audio signal is further enabled by at least one further selection rule using no information on sections of said audio signal preceding said specific section, said at least one further selection rule being applied at least as long as the number of received sections is less than

the number of sections covered by an analysis window, in which signal characteristics are determined for said at least one selection rule.

3. A method according to claim 1 or 2, wherein said at least one selection rule, which is based on signal characteristics that have been determined from an analysis window, comprises a first selection rule, which is based on signal characteristics that have been determined in a shorter analysis window, and a second selection rule, which is based on signal characteristics that have been determined in a longer analysis window, wherein said first selection rule is activated as soon as sufficient sections of said audio signal for said shorter analysis window have been received, and wherein said second selection rule is activated as soon as sufficient sections of said audio signal for said longer analysis window have been received.

4. A method according to claim 3, wherein a respective section of said audio signal corresponds to a respective audio signal frame having a length of 20 ms, wherein said shorter window covers an audio signal frame for which a coding model is to be selected and in addition four preceding audio signal frames, and wherein said longer window covers an audio signal frame for which a coding model is to be selected and in addition sixteen preceding audio signal frames.

5. A method according to one of the preceding claims, wherein said signal characteristics comprise a

standard deviation of energy related values in a respective analysis window.

6. A method according to one of the preceding claims, wherein said first coder mode is an extension mode of an extended adaptive multi-rate wideband codec and enables a coding based on an algebraic code-excited linear prediction coding model and in addition a coding based on a transform coding model, and wherein said second coder mode is an adaptive multi-rate wideband mode of said extended adaptive multi-rate wideband codec and enables a coding based on an algebraic code-excited linear prediction coding model.

7. A method according to any of the preceding claims, wherein said section is a frame or a sub-frame of said audio signal.

8. A module (2,3) for supporting an encoding of an audio signal, said module (2,3) comprising:
   - a first coder mode portion (5) adapted to encode a respective section of an audio signal in a first coder mode;
   - a second coder mode portion (4) adapted to encode a respective section of an audio signal in a second coder mode;
   - switching means (6) for switching between said first coder mode portion (5) and said second coder mode portion (4);
   - comprised by said first coder mode portion (5) an encoding portion (9) which is adapted to encode a respective section of said audio signal based on at least two different coding models; and

- further comprised by said first coder mode portion (5) a selection portion (13,14,15) adapted to apply at least one selection rule for selecting a specific coding model, which coding model is to be used by said encoding portion (9) for encoding said specific section of an audio signal, wherein said at least one selection rule is based on signal characteristics, which have at least partly been determined from an analysis window covering at least one section of an audio signal preceding said specific section, and wherein said selection portion (13,14,15) is adapted to activate said at least one selection rule after a switch by said switching means (6) from said second coder mode portion (4) to said first coder mode portion (5) in response to having received at least as many sections of said audio signal as are covered by said analysis window.

9. A module (2,3) according to claim 8, further comprising a counter (12) adapted to count the number of sections of said audio signal, which are provided to said first coder mode portion (5) after a switch from said second coder mode portion (4) to said first coder mode portion (5).

10. A module (2,3) according to claim 8 or 9, wherein said first coder mode portion (5) further comprises at least one further selection portion (16,17,18), which is adapted to apply at least one further selection rule for selecting a respective coding model, which coding model is to be used by said encoding portion (9) for encoding a specific section of an audio signal, wherein said at least one further

selection rule uses no information on sections of said audio signal preceding said specific section, and wherein said at least one further selection rule is applied after a switch from said second coder mode portion (4) to said first coder mode portion (5) at least as long as the number of sections received by said first coder portion (5) is less than the number of sections covered by an analysis window employed for said at least one selection rule which is based on an analysis of signal characteristics in an analysis window

11. A module (2,3) according to one of claims 8 to 10, wherein said at least one selection portion (13,14,15) comprises a first selection portion (14) adapted to apply a first selection rule which is based on signal characteristics which have been determined in a shorter analysis window and a second selection portion (13) adapted to apply a second selection rule, which is based on signal characteristics that have been determined in a longer analysis window, wherein said first selection rule is activated as soon as sufficient sections of said audio signal for said shorter analysis window have been received by said first coder model portion (5) after a switch from said second coder mode portion (4) to said first coder mode portion (5), and wherein said second selection rule is activated as soon as sufficient sections of said audio signal for said longer analysis window have been received by said first coder model portion (5) after a switch from said second coder mode portion (4) to said first coder mode portion (5).

12. An electronic device (1) supporting an encoding of an audio signal, said electronic device (2,3) comprising:
    - a first coder mode portion (5) adapted to encode a respective section of an audio signal in a first coder mode;
    - a second coder mode portion (4) adapted to encode a respective section of an audio signal in a second coder mode;
    - switching means (6) for switching between said first coder mode portion (5) and said second coder mode portion (4);
    - comprised by said first coder mode portion (5) an encoding portion (9) which is adapted to encode a respective section of said audio signal based on at least two different coding models; and
    - further comprised by said first coder mode portion (5) a selection portion (13,14,15) adapted to apply at least one selection rule for selecting a specific coding model, which coding model is to be used by said encoding portion (9) for encoding said specific section of an audio signal, wherein said at least one selection rule is based on signal characteristics, which have at least partly been determined from an analysis window covering at least one section of an audio signal preceding said specific section, and wherein said selection portion (13,14,15) is adapted to activate said at least one selection rule after a switch by said switching means (6) from said second coder mode portion (4) to said first coder mode portion (5) in response to having received at least as many sections of said audio signal as are covered by said analysis window.

13. An electronic device (1) according to claim 12, further comprising a counter (12) adapted to count the number of sections of said audio signal, which are provided to said first coder mode portion (5) after a switch from said second coder mode portion (4) to said first coder mode portion (5).

14. An electronic device (1) according to claim 12 or 13, wherein said first coder mode portion (5) further comprises at least one further selection portion (16,17,18), which is adapted to apply at least one further selection rule for selecting a respective coding model, which coding model is to be used by said encoding portion (9) for encoding a specific section of an audio signal, wherein said at least one further selection rule uses no information on sections of said audio signal preceding said specific section, and wherein said at least one further selection rule is applied after a switch from said second coder mode portion (4) to said first coder mode portion (5) at least as long as the number of sections received by said first coder portion (5) is less than the number of sections covered by an analysis window employed for said at least one selection rule which is based on an analysis of signal characteristics in an analysis window

15. An electronic device (1) according to one of claims 12 to 14, wherein said at least one selection portion (13,14,15) comprises a first selection portion (14) adapted to apply a first selection rule which is based on signal characteristics which have been determined in a shorter analysis window and a second

selection portion (13) adapted to apply a second selection rule, which is based on signal characteristics that have been determined in a longer analysis window, wherein said first selection rule is activated as soon as sufficient sections of said audio signal for said shorter analysis window have been received by said first coder model portion (5) after a switch from said second coder mode portion (4) to said first coder mode portion (5), and wherein said second selection rule is activated as soon as sufficient sections of said audio signal for said longer analysis window have been received by said first coder model portion (5) after a switch from said second coder mode portion (4) to said first coder mode portion (5).

16. An electronic device (1) according to claim 15, wherein a respective section of said audio signal corresponds to a respective audio signal frame having a length of 20 ms, wherein said shorter window covers an audio signal frame for which a coding model is to be selected and in addition four preceding audio signal frames, and wherein said longer window covers an audio signal frame for which a coding model is to be selected and in addition sixteen preceding audio signal frames.

17. An electronic device (1) according to one of claims 12 to 16, wherein said first coder mode portion (5) further comprises a signal characteristics determination portion (11), which determines signal characteristics of said audio signal in a respective analysis window and which provides said signal characteristics to said selection portion (13,14,15),

said signal characteristics including a standard deviation of energy related values in a respective analysis window.

18. An electronic device (1) according to one of claims 12 to 17, wherein said first coder mode is an extension mode of an extended adaptive multi-rate wideband codec, said encoding portion (9) of said first coder mode portion (5) being adapted to encode sections of an audio signal based on an algebraic code-excited linear prediction coding model and in addition based on a transform coding model, and wherein said second coder mode is an adaptive multi-rate wideband mode of said extended adaptive multi-rate wideband codec, said second coder mode portion (4) being adapted to encode sections of an audio signal based on an algebraic code-excited linear prediction coding model.

19. An audio coding system (1,2) comprising a module (2,3) according to one of claims 8 to 11 and a decoder (20) for decoding audio signals, which have been encoded by said module (2,3).

20. An audio coding system (1,2) according to claim 19, further comprising a first coder mode portion (5) adapted to encode a respective section of an audio signal in a first coder mode.

21. An audio coding system (1,2) according to at least one of the claims 19 and 20, further comprising a second coder mode portion (4) adapted to encode a respective section of an audio signal in a second coder mode.

22. An audio coding system (1,2) according to at least one of the claims 19 to 21, further comprising switching means (6) for switching between said first coder mode portion (5) and said second coder mode portion (4).

23. A software program product, in which a software code for supporting an encoding of an audio signal is stored, wherein at least a first coder mode and a second coder mode are available for encoding a respective section of said audio signal, wherein at least said first coder mode enables a coding of a respective section of said audio signal based on at least two different coding models, and wherein in said first coder mode a selection of a respective coding model for encoding a specific section of an audio signal is enabled by at least one selection rule, which is based on signal characteristics that have been determined from an analysis window, which covers at least one section of said audio signal preceding said specific section, said software code realizing the following step when running in a processing component (3) of an encoder (2):

- activating said at least one selection rule after a switch from said second coder mode to said first coder mode in response to having received at least as many sections of said audio signal as are covered by said analysis window.

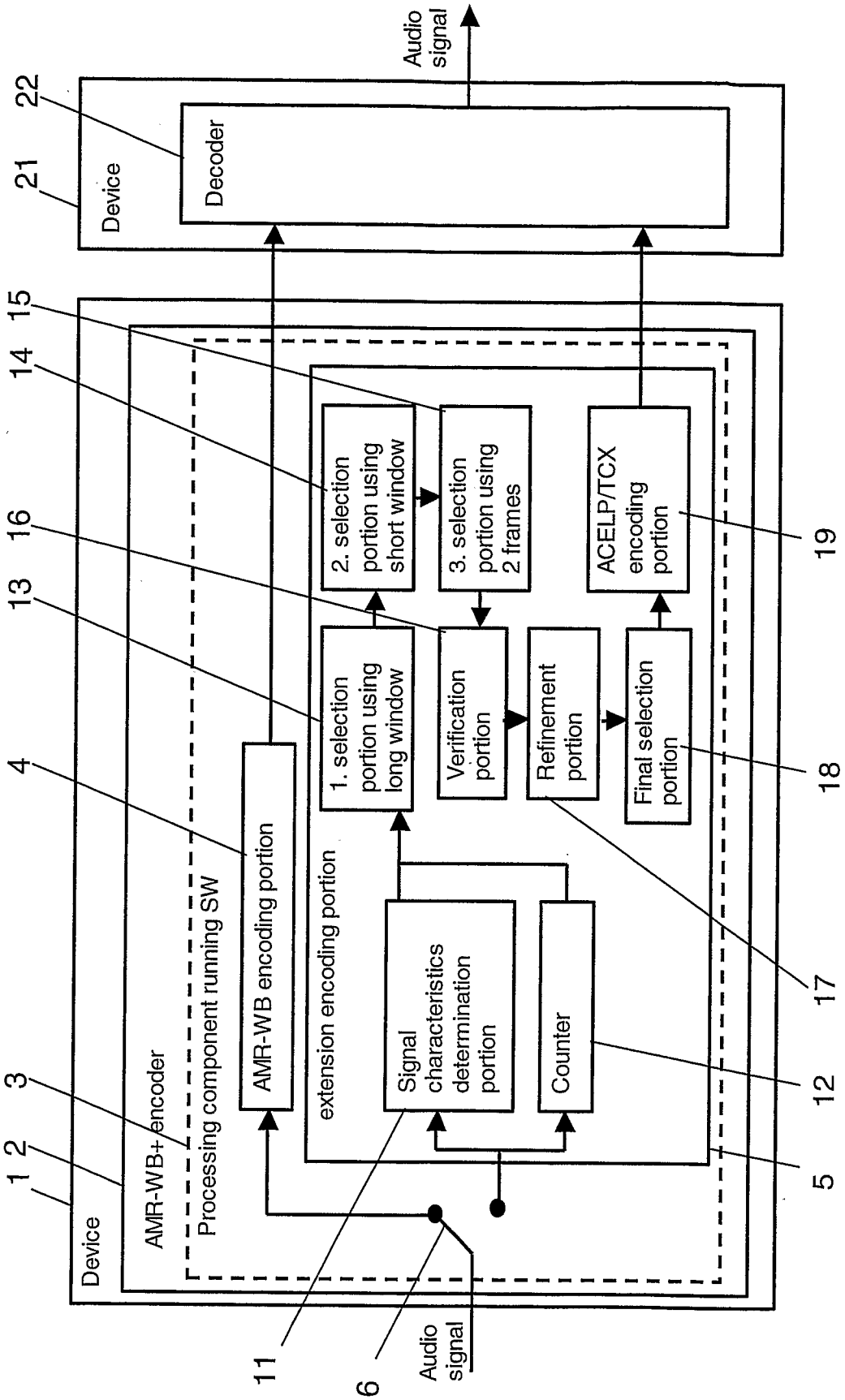

Audio signal
22
21
Device
Decoder
15
14
16
13
4
3
1 2
1
Device
AMR-WB+ encoder
Processing component running SW
AMR-WB encoding portion
extension encoding portion
1. selection portion using long window
2. selection portion using short window
3. selection portion using 2 frames
Verification portion
Refinement portion
Final selection portion
ACELP/TCX encoding portion
Signal characteristics determination portion
Counter
19
18
17
12
5
11
6
Audio signal

Fig. 1

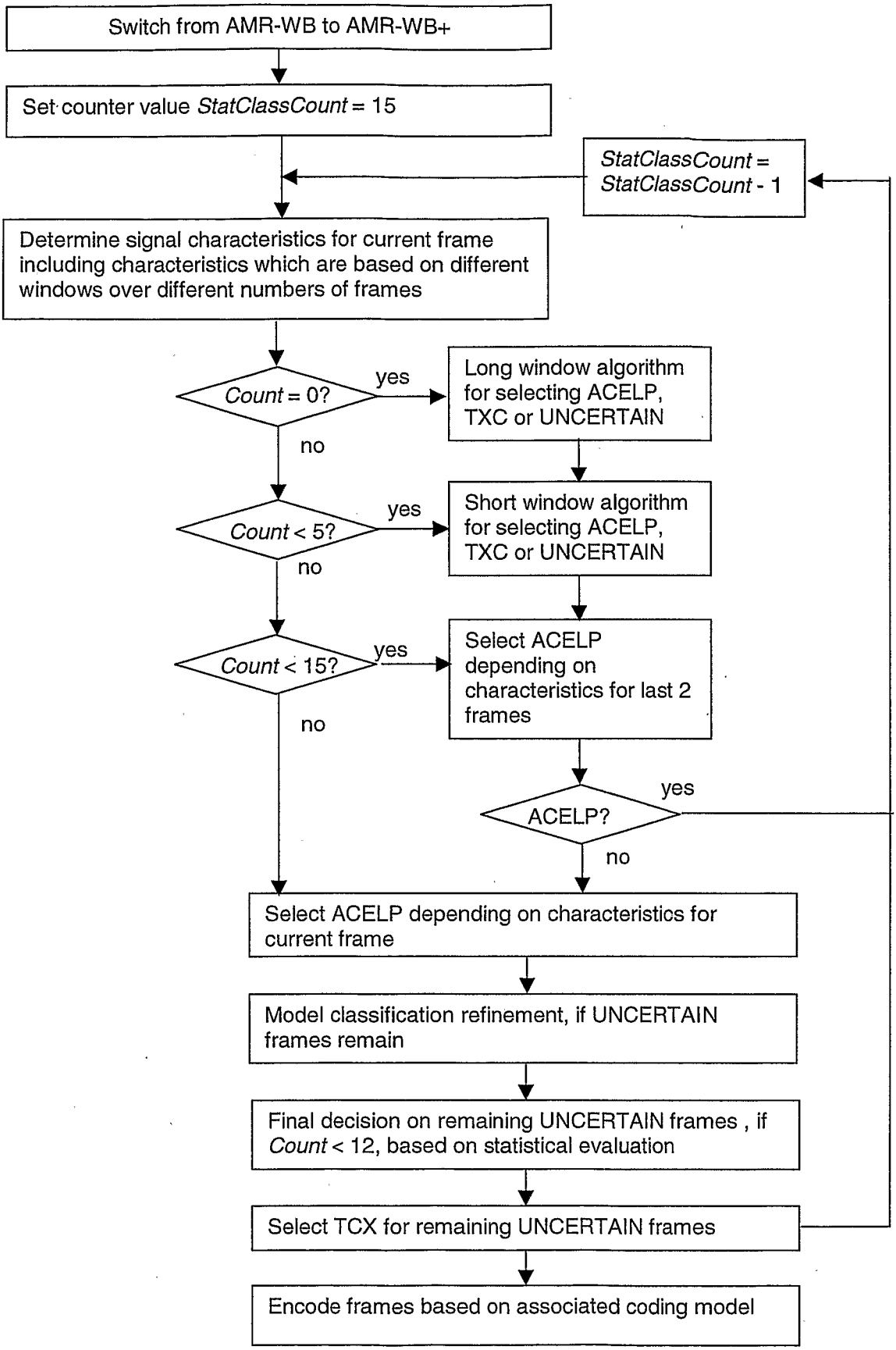
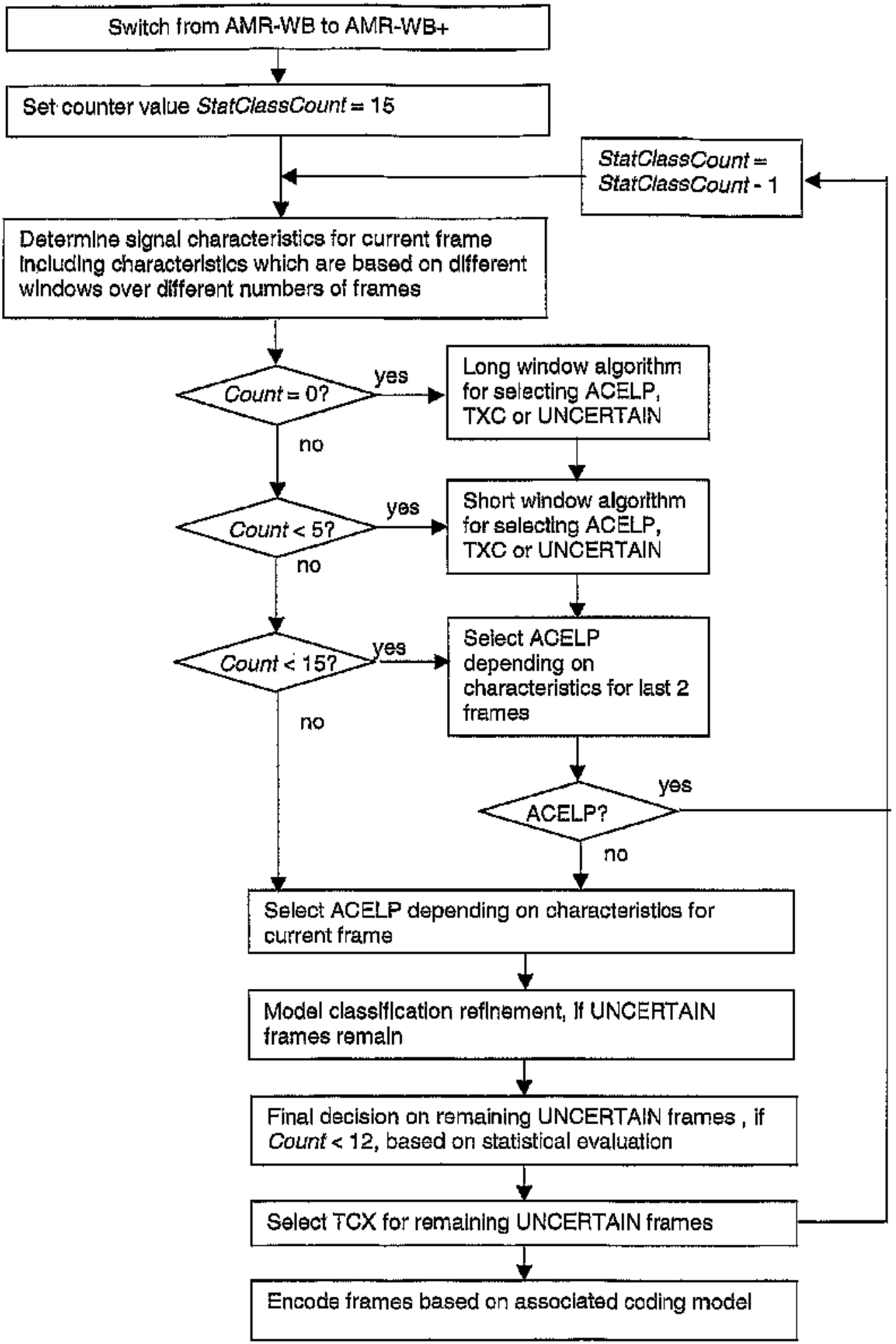

Switch from AMR-WB to AMR-WB+
Set counter value StatClassCount = 15
StatClassCount = StatClassCount - 1
Determine signal characteristics for current frame including characteristics which are based on different windows over different numbers of frames
Count = 0?
yes
Long window algorithm for selecting ACELP, TXC or UNCERTAIN
no
Count < 5?
yes
Short window algorithm for selecting ACELP, TXC or UNCERTAIN
no
Count < 15?
yes
Select ACELP depending on characteristics for last 2 frames
no
ACELP?
yes
no
Select ACELP depending on characteristics for current frame
Model classification refinement, if UNCERTAIN frames remain
Final decision on remaining UNCERTAIN frames , if Count < 12, based on statistical evaluation
Select TCX for remaining UNCERTAIN frames
Encode frames based on associated coding model

Fig. 2


Switch from AMR-WB to AMR-WB+
Set counter value StatClassCount = 15
StatClassCount = StatClassCount - 1
Determine signal characteristics for current frame including characteristics which are based on different windows over different numbers of frames
Count = 0?
yes
no
Long window algorithm for selecting ACELP, TXC or UNCERTAIN
Count < 5?
yes
no
Short window algorithm for selecting ACELP, TXC or UNCERTAIN
Count < 15?
yes
no
Select ACELP depending on characteristics for last 2 frames
ACELP?
yes
no
Select ACELP depending on characteristics for current frame
Model classification refinement, if UNCERTAIN frames remain
Final decision on remaining UNCERTAIN frames , if Count < 12, based on statistical evaluation
Select TCX for remaining UNCERTAIN frames
Encode frames based on associated coding model