

Encoder, decoder and method for encoding and decoding audio content using parameters for enhancing a concealment

The present invention concerns an audio codec, using an encoder and a decoder, in which audio frames that are defective, e.g. lost, corrupted or delayed, are at least partially reconstructed by using an error concealment mechanism. The present invention improves conventional error concealment mechanisms by providing selected error concealment helper parameters within the bitstream, which error concealment helper parameters enhance the decoder-sided concealment.

BACKGROUND OF THE INVENTION

In VoIP systems, packets arrive with different latencies or even in swapped chronological order at the receiver. As each packet is expected at a determined, periodic point of time for decoding at the speech/audio decoder, a so-called de-jitter buffer is needed to remove the time jitter and restore correct order between the packets, if possible.

The availability of a de-jitter buffer enables the usage of channel aware coding, where a partial redundant copy of a current frame is coded on top of a future frame's primary copy within the encoder. If the current frame gets lost or arrives too late at the receiver, its partial redundant copy, which arrives within a later frame, can be used to synthesize the lost frame. The delay (or number of frames) between a primary frame and its partial redundant copy, the so-called FEC offset, as well as the decision, if a partial redundant copy of a particular frame needs to be transmitted at all, can be controlled dynamically at the encoder, depending on the actual available system delay and the frame error rate (FER), i.e. the current channel conditions.

Although this technique requires the total size of the primary frame to be reduced to keep the bitrate constant, it allows for better quality compared to non-channel aware/redundancy based approaches at mid and high FERs.

Networks such as the internet are used for VoIP communication such as conferencing, in addition to sending data. Accordingly, multiple voices or music is encoded into digital data, the data is arranged in packets, and the packets are transmitted to the recipient over a network. VoIP requires that this process happen in real time.

A disadvantage of protocols that permit real time use is that they are unreliable, in that they permit packets to be lost, without retrieving them. When that happens, the voice or audio segments they were carrying are not reconstructed, and the recipient hears annoying gaps in speech or music. These gaps are perceived as reduced quality of service.

- 5 In order to conceal the fact that a packet has been lost, redundancy schemes have been devised. Redundant packets are encoded and transmitted, which repeat aspects of the original data. If a packet is lost, its data is recovered and/or reconstructed from its corresponding redundant packet, which is hopefully not lost. A jitter buffer at the receiving end collects the primary and redundant packets and feeds them to the decoder which plays them out.
- 10 The first media-specific error correction scheme defined for RTP was audio redundancy coding, specified in RFC 2198 [1]. This was designed for voice teleconferences. Each packet contains both an original frame of audio data and a redundant copy of a preceding frame, in a more heavily compressed format.

- Packet-based traffic can be subject to high packet loss ratios, jitter and reordering. Forward error correction (FEC) is one technique for addressing the problem of lost packets. Generally, FEC involves transmitting redundant information along with the coded speech. The decoder attempts to use the redundant information to reconstruct lost packets. Media-independent FEC techniques add redundant information based on the bits within the audio stream (independent of higher-level knowledge of the characteristics of the speech stream). On the other hand,
- 15 media-dependent FEC techniques add redundant information based on the characteristics of the speech stream.
- 20

The granted patent US 6,757,654 [2] describes an improved FEC technique for coding speech data. US 6,757,654 discloses:

- "[This technique consist of] an encoder module primary-encodes an input speech signal using a primary synthesis model to produce primary-encoded data, and redundant-encodes the input speech signal using a redundant synthesis model to produce redundant-encoded data. A packetizer combines the primary-encoded data and the redundant-encoded data into a series of packets and transmits the packets over a packet-based network, such as an Internet Protocol (IP) network. A decoding module primary-decodes the packets using the primary synthesis model, and redundant-decodes the packets using the redundant synthesis model.
- 25
- 30 The technique provides interaction between the primary synthesis model and the redundant synthesis model during and after decoding to improve the quality of the synthesized output speech signal. Such "interaction," for instance, may take the form of updating states in one model using the other model.

Further, the present technique takes advantage of the FEC-staggered coupling of primary and redundant frames (i.e., the coupling of primary data for frame n with redundant data for frame $n-1$) to provide look-ahead processing at the encoder module and the decoder module. The look-ahead processing supplements the available information regarding the speech signal, and thus improves the quality of the output synthesized speech.

The interactive cooperation of both models to code speech signals greatly expands the use of redundant coding heretofore contemplated by conventional systems."

The conference paper [3] presents a joint playout buffer and Forward Error Correction (FEC) adjustment scheme for Internet Telephony, which incorporates the impact of end-to-end delay on the perceived audio quality. Conference paper [3] represents the perceived audio quality as a function of both the end-to-end delay and the distortion of the voice signal. A joint rate/error/playout delay control algorithm is developed that optimizes this measure of quality.

As said in [3], Media specific FEC is used by most audio conferencing tools. The principle of the signal processing FEC is to transmit each segment of audio, encoded with different quality coders, in multiple packets. When a packet is lost, another packet containing the same segment (maybe encoded differently) can be able to cover the loss.

All the state of the art is based on redundancy, which means sending a really low bitrate version of the current frame with a later frame. Although redundant audio encoding can provide exact repair (if the redundant copy is identical to the primary) it is more likely that a lower bitrate will be used and hence lower quality will be achieved. In the context of advance Speech and audio coding the data rate is getting big for each frame and transmitting a really low bitrate version of it leads to relatively poor quality.

Thus, it is desired to improve existing error concealment mechanisms.

SUMMARY OF THE INVENTION

Embodiments of the present invention create an improved solution by providing an encoder having at least the features of independent claim 1, a decoder having at least the features of independent claim 22, a system according to claim 42, a method for encoding comprising at least the features of independent claim 43, a method for decoding comprising at least the features of independent claim 44, a computer program according to claim 45, an encoder having at least the features of claim 46, a decoder having at least the features of claim 55, an apparatus according to at least one of independent claims 62 to 64, a switched coder or

decoder having at least the features of independent claim 65, a Transform Domain Coder or decoder having at least the features of independent claim 66 and a method or computer program according to independent claim 68.

According to claim 1, it is proposed to provide an encoder for coding speech-like content and/or general audio content, wherein the encoder is configured to embed, at least in some frames, parameters in a bitstream, which parameters enhance a concealment in case an original frame is lost, corrupted or delayed. Even though standard concealment mechanisms may be used for a lost frame, the parameters that are embedded in the frames will be used to enhance this concealment. Accordingly, this invention proposes to not have a partial copy that is just a low bitrate version of the primary, but to transmit some selected parameters only that will enhance a concealment. Therefore the decoder may work different from decoders as proposed in the state of the art.

It has been found that the provision of some selected parameters which enhance the error concealment (e.g. which define characteristics of a lost frame which would otherwise need to be estimated on the basis of a previous frame preceding a defective frame that has been lost, corrupted or delayed) brings along a good error concealment (of a defective frame) while keeping a required bitrate low.

Worded differently, the transmission of the parameters which enhance the concealment makes it possible to reconstruct a defective frame on the basis of information about previously decoded frames, wherein most of the information of the concealed frame is derived from one or more frames preceding (or following) the defective frame, but wherein one or more of the most relevant characteristics of the defective frame (or one or more of the most important parameters of the error concealment), which would normally need to be derived from the preceding or following correctly coded frames, are represented in a comparably accurate manner by the parameters which enhance the concealment.

Worded yet differently, the embedded parameters for enhancing the error concealment are preferably insufficient for a reconstruction of a defective frame in that they do not contain all required types of information but support an error concealment in that the most important types of information are provided by the parameters while other types of information for the concealment must be derived from previously decoded frames at the decoder side.

Accordingly, a good compromise between error concealment quality and bitrate is achieved.

In an embodiment, the encoder may be configured to create a primary frame and a so-called "partial copy", wherein the "partial copy" is not a low bitrate version of the primary frame but wherein the "partial copy" contains the parameters (e.g. some of the most relevant parameters

required for concealing if the frame under consideration is defective). In other words, the “partial copy” as used herein is not a low bitrate representation of the (original) audio content being embedded as redundant information to the bitstream, and which may later be used to fully synthesize the output signal. Instead, it is the inventive concept to embed some parameter data, namely the aforementioned parameters which enhance the concealment at the decoder side, if said parameter data is available. When using this information, the decoder has to be in a concealment mode. Accordingly, the decoder will decode the “partial copy” of a defective, i.e. lost, corrupted or delayed frame (possibly available due to a de-jitter buffer delay) and use said decoded parameters to assist the concealment routine at the decoder side. Thus, the size that may be needed to encode a partial copy, comprising only one or more parameters, can be reduced when compared to the size needed to encode a redundant copy by redundant-encoding the content of an entire primary frame (e.g. at a reduced bitrate), whereas would generally also be possible to use the same bitrate or a higher bitrate for encoding a partial copy. However, the inventive concept, i.e. enhancing a concealment by error concealment helper parameters, provides for a better quality compared to conventional decoding of a low bitrate version of the respective primary frame.

In an embodiment, the encoder may be configured to delay the parameters by some time and to embed the parameters in a packet which is encoded and sent later in time. In other words, the encoder first sends the primary frame in a first packet. With a certain time delay, the encoder then sends the “partial copy” in another packet which is sent later than the first packet. Accordingly, the encoder still quantizes the parameters but adds them to the bitstream in a later packet. Thus, even when a primary frame is unavailable or defective, e.g. lost, corrupted or delayed, its content may still be correctly reconstructed (or at least approximated without severe artefacts) at the decoder side by means of a concealment with the help of the parameters that have been sent later and which might therefore be available at the decoder.

In an embodiment, the encoder may be configured to reduce a primary frame bitrate, wherein the primary frame bitrate reduction and a partial copy frame coding mechanism together determine a bitrate allocation between the primary frames and partial copy frames to be included within a constant total bitrate. Thus, the encoder provides for a constant total bitrate when sending primary frames and partial copy frames, while at the same time providing good audio quality with low perceptual impact.

In an embodiment, the encoder may be configured to create a primary frame of one of the speech-like content type and the general audio content type in combination with a partial copy of the other one of the speech-like content type and the general audio content type. Thus, the encoder is versatile as it can handle different types of audio content separately or in combination with each other. This is particularly useful as the encoder is thus adapted to

combine, for example, an ACELP primary frame with a TCX partial redundant copy, or vice versa.

In an embodiment, the encoder may be part of a codec using a TCX coding scheme. According to this embodiment, the encoder preferably uses TCX coding for efficiently encoding general audio content, music, background noise, or the like. The encoder can reliably determine and transmit TCX specific parameters that can be used for TCX concealment at the decoder side when the partially redundant frame may, for example, not include any encoded spectral values and may therefore by itself not be sufficient to reconstruct the defective frame.

In an embodiment, the encoder may be configured to detect whether the frame contains a noisy, or Noise-like, audio signal or whether the frame contains a noise floor with sharp spectral lines that are stationary over a period of time, and to embed, based on the detection, the parameters into a TCX frame. Thus, a decision on the current signal characteristic can already be made at the encoder side such that specific parameters for those signals are encoded and sent to the decoder for enhancing the concealment.

In an embodiment, the parameters may comprise ISF or LSF parameters, in particular predictively coded ISF or LSF parameters. ISF and LSF parameter representation is used for quantization and coding of LPC parameters. In a TCX coding scheme the LPC is used to represent the masking threshold. This is an important parameter and very helpful to have available correctly on decoder side in case of a frame loss. Especially if the ISF/LSFs are coded predictively the concealment quality will improve by having this info available during concealment, because the predictor states on decoder side will stay correct, i.e. in sync to the encoder, and this will lead to a quick recovery of an unavailable primary frame.

In an embodiment, the parameters may comprise signal classification parameters. Signal classification is used for signaling the content types: UNVOICED, UNVOICED TRANSITION, VOICED TRANSITION, VOICED and ONSET. Typically this type of classification is used in speech coding and indicating if tonal/predictive components are present in the signal or if the tonal/predictive components are changing. Having this information available on the decoder side during concealment may help to determine the predictability of the signal and thus it can help adjusting the amplitude fade-out speed, the interpolation speed of the LPC parameters.

In an embodiment, the parameters may comprise a TCX global gain or a TCX global level. The global gain may be transmitted to easily set the energy of the concealed frame to the correct (encoder determined level) in case it is available.

In an embodiment, the parameters may comprise at least one of a window information and a spectral peak position. Having this information available already at the encoder side is useful for selectively transmitting those parameters to the decoder for concealment.

In an embodiment, the encoder may be part of a switched codec, wherein the switched codec consists of at least two core coding schemes, wherein a first core coding scheme uses ACELP and a second core coding scheme uses TCX. For example, the encoder uses ACELP for coding speech-like audio content and TCX for coding general audio content. Thus, using several coding schemes for encoding audio content renders the encoder versatile. Furthermore, the encoder provides good results by using a signal specific coding scheme for each signal.

In an embodiment, the encoder may be configured to not put a "partial copy" on top of a TCX frame after a switch when there is a first TCX frame after an ACELP frame. For example, the provision of parameters enhancing a concealment may be selectively omitted in this case. If the first TCX frame is lost, it is not possible to conceal in TCX mode. Thus, ACELP concealment will be used instead. In this case, TCX partial copies alone will not be sufficient to fully synthesize the frame, the decoder needs to be in concealment mode and may be supported by partial copies. Thus, as concealment needs a preceding frame for extrapolating the signal content, it is preferable in this case to use ACELP concealment (as the previous frame was ACELP) which would make a TCX partial copy less useful. As the encoder is configured to detect a switch and to selectively, i.e. depending on a switch event, provide a certain type of partial copy, the concealment at the decoder side will provide a good result.

In an embodiment, the encoder may be configured to analyze the signal before encoding and to turn off the partial copy usage (e.g. not provide any parameters) or to provide a reduced partial copy (e.g. provide less parameters than in a normal case) based on the analyzed signal. For example, if a signal could be concealed satisfyingly well without the help of additional partial copy info within the decoder, but the clean channel performance suffers because of reduced primary frame, partial copy usage can be turned off or a specifically reduced partial copy can be used within the encoder. Thus, the encoder is adapted to selectively provide a partial copy, i.e. to provide a partial copy only if concealment parameters are needed at the decoder side for reconstructing audio content of an unavailable primary frame. Furthermore, the bandwidth-usage of the primary frame transmission can be optimized.

In an embodiment, the encoder may be configured to choose between multiple partial copy modes which use different amounts of information and/or different parameter sets, wherein the selection of the partial copy mode is based on parameters (e.g. parameters describing the signal to be encoded). Thus, the encoder can selectively choose a certain partial copy mode

for providing a partial copy that is well suited for concealing a certain unavailable primary frame at the decoder side. The selection between multiple partial copy modes is based on various parameters, such as the current and/or previous frame's signal characteristics, including pitch stability, LTP pitch, LTP gain, the temporal trend of the signal, the mode of the last two frames and a frame class.

In an embodiment, at least one of the multiple partial copy modes may be a frequency domain concealment mode. This mode can selectively be chosen by the encoder for providing a partial copy comprising certain parameters that are well suited for providing, at the decoder side, a good concealment result of an unavailable primary frame containing a frequency domain signal.

In an embodiment, at least two of the multiple partial copy modes may be different time domain concealment modes. For example, a first partial copy mode could be selected if the respective time domain signal comprises at least a certain characteristic. Otherwise, if the time domain signal does not comprise this certain characteristic, or if the time domain signal comprises a different signal characteristic, the second partial copy mode is chosen. Thus, the encoder provides for a signal specific selection of the parameters contained in a partial copy.

In an embodiment, one of the at least two time domain concealment modes can be selected if a frame contains a transient or if a global gain of the frame is lower (e.g. at least by a predefined amount) than a global gain of a previous frame. Thus, the encoder selectively chooses a mode for providing parameters which are used, at the decoder side, for enhancing a concealment of a defective or unavailable primary frame, even if this defective or unavailable primary frame's signal characteristics deviate to a certain extent from the previous frame's signal characteristic.

In an embodiment, the encoder may be configured to send (as a parameter for enhancing a concealment) a LTP lag if LTP data is present. Thus, the encoder selectively provides parameters used, at the decoder side, for Long Term Prediction decoding.

In an embodiment, the encoder may be configured to send (as a parameter for enhancing a concealment) a classifier information. Signal classification is used for signaling the content types: UNVOICED, UNVOICED TRANSITION, VOICED TRANSITION, VOICED and ONSET. Typically, this type of classification is used in speech coding and indicating if tonal/predictive components are present in the signal or if the tonal/predictive components are changing. Having this information available on the decoder side (sent by the encoder) during concealment may help to determine the predictability of the signal and thus it can help adjusting the amplitude fade-out speed and/or the interpolation speed of LPC parameters and it can

control possible usage of high- or low pass filtering of voiced or unvoiced excitation signals (e.g. for de-noising).

In an embodiment, the encoder may be configured to send (as a parameter for enhancing a concealment) at least one of LPC parameters, LTP Gain, Noise Level and Pulse Position.

5 Thus, the encoder transmits certain parameters that are well suited for concealing, at the decoder side, the content of a defective or unavailable primary frame (i.e. to enhance the concealment).

Another embodiment provides a decoder for decoding speech-like content and/or general audio content, wherein the decoder is configured to use parameters which are sent later in time to enhance a concealment in case an original frame is lost, corrupted or delayed. Accordingly, at the receiver (or decoder), the parameters which are sent later in time can be used for enhancing an error concealment at the decoder side and thus recreating a signal (e.g. a concealed signal which avoids severe artefacts) if the original frame is defective, e.g. lost, corrupted or delayed. Thus, the inventive concept can reliably reconstruct unavailable audio content by using parameters enhancing a concealment while efficiently using a given bandwidth.

For example, the parameters which are sent to enhance the concealment (and which are evaluated) by the audio decoder may comprise one or more of the most important information types which are required in a concealment of a defective frame by a concealment unit of the decoder. However, the parameters are typically chosen such that the parameters alone are insufficient to perform a full error concealment. Rather, for actually performing the error concealment, the concealment unit of the decoder typically obtains additional information types, for example, on the basis of previously (or subsequently) decoded frames. Thus, the parameters which are sent later in time merely enhance the concealment, but they do not constitute a full concealment information.

Accordingly, the usage of the parameters which are sent later in time allows to have a precise information about the most important concealment parameters available at the audio decoder with only small bitrate effort, while additional information required for providing a concealed frame is generated by the audio decoder itself, for example on the basis of one or more previously (or subsequently) decoded frames using extrapolation or interpolation.

In an embodiment, the decoder may be configured to receive a primary frame and a "partial copy", wherein the "partial copy" is not a low bitrate version of the primary frame but wherein the "partial copy" contains the parameters to enhance a concealment. As the "partial copy" contains these parameters, the bandwidth used for the transmission of these parameters is

even lower as compared to the bandwidth used for transmitting a low bitrate version of the primary frame.

In an embodiment, the parameters are contained in a partial copy and the decoder is configured to receive from a de-jitter buffer the partial copy of a currently lost frame if it is available. A de-jitter buffer further improves the inventive concept as it is able to provide a jitter delay, wherein a certain number of frames can be buffered. Thus, frames that arrive at the decoder in a wrong chronological order (i.e. a first frame that has been sent at the encoder side prior to a second frame arrives later at the decoder side than the second frame, even though the first frame is expected to arrive earlier at the decoder side than the second frame) can be buffered and provided in the correct chronological order. This is particularly useful if a frame is delayed.

In an embodiment, the decoder may be configured to receive a primary frame of one of the speech-like content type and the general audio content type in combination with a partial copy of the other one of the speech-like content type and the general audio content type. Thus, the decoder is versatile as it can handle different types of audio content separately or in combination with each other. This is particularly useful as the decoder is thus adapted to extract, for example, a TCX partial redundant copy that has been transported on top of an ACELP primary frame, or vice versa.

In an embodiment, the decoder may be part of a codec using a TCX codec scheme. According to this embodiment, the decoder preferably uses TCX decoding for efficiently decoding general audio content, music, background noise, or the like. The decoder can reliably extract TCX specific parameters (to enhance a concealment) from a partial copy for enhancing a TCX concealment.

In an embodiment, the parameters may comprise ISF or LSF parameters, in particular predictively coded ISF or LSF parameters. ISF and LSF parameter representation is used for quantization and coding of LPC parameters. In a TCX coding scheme the LPC is used to represent the masking threshold. This is an important parameter and very helpful to have available correctly on decoder side in case of a frame loss. Especially if the ISF/LSFs are coded predictively the concealment quality will improve by having this info available during concealment, because the predictor states on decoder side will stay correct, i.e. in sync to the encoder, and this will lead to a quick recovery of an unavailable primary frame.

In an embodiment, the parameters may comprise signal classification parameters. Signal classification is used for signaling the content types: UNVOICED, UNVOICED TRANSITION, VOICED TRANSITION, VOICED and ONSET. Typically this type of classification is used in

speech coding and indicating if tonal/predictive components are present in the signal or if the tonal/predictive components are changing. Having this information available on the decoder side during concealment may help to determine the predictability of the signal and thus it can help adjusting the amplitude fade-out speed, the interpolation speed of the LPC parameters.

- 5 In an embodiment, the parameters may comprise a TCX global gain or a TCX global level. The global gain may be transmitted to easily set the energy of the concealed frame to the correct (encoder determined) level in case it is available.

In an embodiment, the parameters may comprise at least one of a window information and a spectral peak position. Having this information available at the decoder side is useful for
10 selectively enhancing the concealment.

In an embodiment, the decoder may be part of a switched codec, wherein the switched codec consists of at least two core coding schemes, wherein a first core coding scheme uses ACELP and a second core coding scheme uses TCX. For example, the decoder uses an ACELP decoding scheme for decoding speech-like audio content and a TCX decoding scheme for
15 decoding general audio content. Thus, using several decoding schemes for decoding different audio content renders the decoder versatile.

In an embodiment, the decoder may be configured to use, after a switch, ACELP concealment in the case that a first TCX frame after an ACELP frame is not available to the decoder. If the first TCX frame is defective, i.e. lost, corrupted or delayed, it is not possible to conceal in TCX
20 mode. Thus, ACELP concealment will be used instead. In this case, TCX partial copies alone will not be sufficient to fully synthesize the frame, the decoder needs to be in concealment mode and may be supported by partial copies. As concealment needs a preceding frame for extrapolating the signal content, it is preferable in this case to use ACELP concealment (as the previous frame was ACELP) which would make a TCX partial copy less useful.

25 In an embodiment, the decoder may be configured to choose between multiple partial copy modes or concealment modes which use different amounts of information and/or different parameter sets among a plurality of several modes available at the decoder. In an embodiment, the decoder chooses the concealment mode if the decoder does not get the respective mode, i.e. if it cannot determine or otherwise retrieve it, from the partial copy. Otherwise, the
30 concealment mode is dictated by the available partial copy, wherein it is the encoder that makes the decision then. Accordingly, the decoder uses the respectively coded different amounts of information and/or different parameter sets directly from the bitstream sent at the encoder side. Thus, the decoder can apply a well-suited concealment mode based on the partial copy mode, wherein there is more supporting (enhancement) information (i.e.

parameters) in one mode and less in another mode. In other words, in CA-mode, the encoder decides on the appropriate concealment mode and prepares the partial copy accordingly. If a partial copy is available to the decoder and it should be used for enhancing the concealment, the decoder must stick to the decision made by the encoder, otherwise the info within the partial copy cannot be exploited properly. The decoder only decides itself on the concealment mode, if no partial copy is available or if the partial copy is not and/or should not be used for other reasons.

In an embodiment, at least one of the multiple partial copy modes may be a frequency domain concealment mode. This mode can selectively be chosen by the decoder for using a partial copy comprising certain parameters that are well suited for providing a good concealment result of an unavailable primary frame containing a frequency domain signal.

In an embodiment, at least two of the multiple partial copy modes may be different time domain concealment modes. For example, a first partial copy contains parameters of a respective time domain signal comprising at least a certain characteristic, while a second partial copy contains parameters of a respective time domain signal comprising a different signal characteristic. One of these two time domain modes can selectively be chosen by the decoder for using a partial copy comprising certain parameters that are well suited for providing a good concealment result of an unavailable primary frame containing a time domain signal.

In an embodiment, the decoder may be configured to receive a LTP lag if LTP data is present in the corresponding primary frame. Thus, the decoder is enabled to reconstruct the content of an unavailable primary frame by Long Term Prediction decoding thereby using the LTP parameters that have been received in a partial copy.

In an embodiment, the decoder may be configured to receive a classifier information. Signal classification is used for signaling the content types: UNVOICED, UNVOICED TRANSITION, VOICED TRANSITION, VOICED and ONSET. Typically, this type of classification is used in speech coding and indicating if tonal/predictive components are present in the signal or if the tonal/predictive components are changing. Having this information available on the decoder side (sent by the encoder) during concealment may help to determine the predictability of the signal and thus it can help adjusting the amplitude fade-out speed, the interpolation speed of LPC parameters and it can control possible usage of high- or low pass filtering of voiced or unvoiced excitation signals (e.g. for de-noising).

In an embodiment, the decoder may be configured to receive (as the parameters for enhancing a concealment) at least one of LPC parameters, LTP Gain, Noise Level and Pulse Position.

Thus, the decoder is enabled to reconstruct the content of an unavailable primary frame by using at least one of these parameters that have been received in a partial copy.

In an embodiment, the decoder may be configured to decrease a pitch gain and a code gain with two different factors in dependence on a concealment mode. This serves to avoid having a long stationary signal whenever the original signal was more transient like.

In an embodiment, a first factor to decrease a pitch gain and a code gain is 0.4 and a second factor is 0.7. These two factors are particularly efficient in order to avoid having a long stationary signal whenever the original signal was more transient like.

In an embodiment, the decoder may be configured to not take into account a pitch decoded from the partial copy if the previous primary frame is lost, and wherein the decoder is configured to fix, i.e. to adjust, the pitch to a predicted pitch for the following lost primary frame instead of using the pitch transmitted. Accordingly, the pitch decoded from the partial copy shall not be taken into account if the previous frame is lost, because the pitch sent in the bitstream was computed on the encoder side based on the ground truth, but if the previous frame is lost, the synthesis of the previously lost and concealed synthesis might be really different to the encoder ground truth. So it is better in general to not risk relying on the synchronicity of en-/decoder in case of multiple frame loss and fix the pitch to the predicted pitch for the following lost frame instead of using the pitch transmitted.

Another embodiment creates a method for encoding speech-like content and/or general audio content, the method comprising the step of embedding, at least in some frames, parameters in a bitstream, which parameters enhance a concealment in case an original frame is lost, corrupted or delayed. Even though standard concealment mechanisms may be used for a defective, i.e. lost, corrupted or delayed frame, the parameters that are embedded in the frames are used by the inventive method in order to enhance this concealment (and the bitstream parameters may replace parameters which are conventionally derived at the decoder side). Accordingly, this invention proposes to not have a partial copy that is just a low bitrate version of the primary, but to transmit parameters that will enhance a concealment (but which typically do not constitute a full error concealment information). Therefore the decoder may be somewhat modified when compared to the state of the art.

Another embodiment creates a method for decoding speech-like content and/or general audio content, the method comprising the step of using parameters which are sent later in time to enhance a concealment in case an original frame is lost, corrupted or delayed. Accordingly, at the receiver, the parameters which are sent later in time can be used for enhancing an error concealment at the decoder side and thus recreating a signal if the original frame is defective,

i.e. lost, corrupted or delayed. Thus, by using the inventive method, defective, corrupted or unavailable audio content can reliably be reconstructed (at least partially) by using parameters instead of an entire redundant coded frame.

Another embodiment creates an encoder for coding audio content, wherein the encoder is configured to provide a primary encoded representation of a current frame and an encoded representation of at least one error concealment parameter for enhancing a decoder-sided error concealment of the current frame, wherein the encoder is configured to select the at least one concealment parameter based on (or in dependence on) one or more parameters representing a signal characteristic of the audio content contained in the current frame. For example and therefore not limiting, the parameters representing a signal characteristic may be chosen from at least the current and previous frame's signal characteristics, including pitch stability, LTP pitch, LTP gain, the temporal trend of the signal, the mode of the last two frames and a frame class. Based on these signal characteristic parameters, the encoder selectively chooses one or more concealment parameters which are well suited for an error concealment at the decoder side. These error concealment parameters are separately encoded, i.e. separately from the primary encoded representation of the signal to be transmitted. Thus, the decoder can reconstruct the signal from these error concealment parameters by using an error concealment, even if the primary encoded representation of that signal is lost, corrupted or delayed. Accordingly, at least in some frames (or packets) error concealment parameters (also designated as redundant coding parameters) are embedded in the bitstream and transmitted to the decoder side. Thus, it is not necessary to provide a "partial copy" of the entire signal, which is usually encoded at a lower bitrate and may therefore comprise a lower quality. Thus, the present invention provides for an improved concept to conceal defective, e.g. lost, corrupted or delayed frames by means of selected error concealment parameters that are already selected (for example in accordance with signal characteristics) at the encoder side and embedded in the bitstream. Thus, the invention keeps within a given bandwidth while at the same time preserving a good quality of the transmitted signal even if a portion (e.g. a frame) of this signal is reconstructed by concealment at the decoder side.

In an embodiment, the decoder-sided error concealment is an extrapolation-based error concealment. Accordingly, the concealment routine may use extrapolation in order to estimate or predict the future signal characteristics, which may further help and assist the concealment of defective primary frames.

In an embodiment, the encoder may be configured to combine the encoded representation of the at least one error concealment parameter of the current frame with a primary encoded representation of a future frame into a transport packet such that the encoded representation of the at least one error concealment parameter of the current frame is sent with a time delay

relative to the primary encoded representation of the current frame. In other words, the encoder first sends a primary frame (i.e. the primary encoded representation of a frame) in a first packet. With a certain time delay, the encoder then sends the "partial copy" (i.e. the encoded representation of the at least one error concealment parameter) in another packet which is sent later than the first packet. Accordingly, the encoder still quantizes the parameters but adds them to the bitstream in a later packet. Thus, the invention is particularly useful in packet-based networks, such as Voice-over-IP (VoIP), Voice-over-LTE (VoLTE) or the like. While the primary encoded representation of a frame may have already been transmitted to the decoder side, its corresponding error concealment parameters will be sent with one of the following transport packets. Thus, if the packet containing the primary encoded representation is lost, corrupted or delayed, the packet containing the error concealment parameters may, however, correctly arrive at the decoder side, as it has been sent later in time. Furthermore, by combining into one packet these error concealment parameters with a primary encoded representation of another frame, bandwidth can be efficiently used.

In an embodiment, the encoder may be configured to selectively choose between at least two modes for providing an encoded representation of error concealment parameters. Thus, the encoder is versatile as it provides different modes for handling different signals that may have different signal characteristics, wherein different sets of error concealment parameters may be provided in the different modes. As these two modes are used for providing an encoded representation of at least one error concealment parameter, these at least two modes are also referred to as partial copy modes.

In an embodiment, the encoder's selection of a mode for providing an encoded representation of the at least one error concealment parameter may be based on one or more parameters which comprise at least one of a frame class, a LTP pitch, a LTP gain and a mode for providing an encoded representation of the at least one error concealment parameter of one or more preceding frames. These parameters are well suited for deciding about a mode for an error concealment at the decoder side.

In an embodiment, at least one of the modes for providing an encoded representation of the at least one error concealment parameter may be a time domain concealment mode such that the encoded representation of the at least one error concealment parameter comprises one or more of a TCX LTP lag and a classifier information. For example, a first mode which is a time domain concealment mode could be selected if a time domain signal is present comprising at least a certain characteristic. Otherwise, if the time domain signal does not comprise this certain characteristic, or if the time domain signal comprises a different signal characteristic, a second mode is chosen. Thus, the encoder provides for a signal specific selection of the error concealment parameters.

In an embodiment, at least one of the modes for providing an encoded representation of the at least one error concealment parameter may be a time domain concealment mode that is selected if the audio content contained in the current frame contains a transient or if the global gain of the audio content contained in the current frame is lower than the global gain of the preceding frame. Thus, the encoder selectively chooses a mode for providing error concealment parameters which are used, at the decoder side, for concealing an unavailable primary encoded representation, even if this unavailable primary frame's signal characteristics deviate to a certain extent from the preceding frame's signal characteristic.

In an embodiment, at least one of the modes for providing an encoded representation of the at least one error concealment parameter may be a frequency domain concealment mode such that the encoded representation of the at least one error concealment parameter comprises one or more of an LSF parameter, a TCX global gain and a classifier information. This mode can selectively be chosen by the encoder for providing an encoded representation of the at least one error concealment parameter which parameter is well suited for providing, at the decoder side, a good concealment result of an unavailable primary encoded representation containing a frequency domain signal.

In an embodiment, the encoder may use at least a TCX coding scheme. According to this embodiment, the encoder preferably uses TCX coding for efficiently encoding general audio content, music, background noise, or the like. Thus, the encoder can reliably determine and transmit TCX specific parameters that can be used for TCX concealment at the decoder side.

An embodiment creates a decoder for decoding audio content, wherein the decoder is configured to receive a primary encoded representation of a current frame and/or an encoded representation of at least one error concealment parameter for enhancing a decoder-sided error concealment of the current frame, wherein the decoder is configured to use the error concealment for at least partly reconstructing the audio content of the current frame by using the at least one error concealment parameter in the case that the primary encoded representation of the current frame is lost, corrupted or delayed. Generally, the decoder is able to receive a bitstream that could be either a single primary frame (i.e. primary encoded representation of a current frame) without any side data (i.e. at least one error concealment parameter) if the encoder decided to not send any side data for the specific past frame, or a primary frame (i.e. primary encoded representation of a current frame) and at least one or more error concealment parameters. Thus, the decoder can at least partially reconstruct a signal using these one or more error concealment parameters by using an error concealment, even if the primary encoded representation of that signal is defective, e.g. lost, corrupted or delayed. Accordingly, at least in some frames error concealment parameters (redundant coding parameters) are embedded in the bitstream and transmitted to the decoder side. Thus,

it is not necessary to provide a partial copy of the entire signal, which is usually encoded at a lower bitrate and may therefore comprise a lower quality. Thus, the present invention provides for an improved concept to conceal defective, e.g. lost, corrupted or delayed frames by using selected error concealment parameters that are already selected at the encoder side, embedded in the bitstream and transmitted to the decoder side, when a concealment that uses information obtained on the basis of one or more previously decoded frames is "guided" (e.g. enhanced or improved) using the received error concealment parameters. Thus, the inventive concept keeps within a given bandwidth (by using an extrapolation-based error concealment which does not require that all error concealment information is transmitted from an encoder to a decoder) while at the same time preserving a good quality of the decoded signal (by enhancing the error concealment using the error concealment parameters) even if the signal is reconstructed by concealment at the decoder side.

In an embodiment, the decoder-sided error concealment is an extrapolation-based error concealment. Accordingly, the concealment routine provided at the decoder side may use extrapolation in order to estimate or predict the future signal characteristics, which may further help and assist the concealment of defective primary frames.

In an embodiment, the decoder may be configured to extract the error concealment parameter of a current frame from a packet that is separated from a packet in which the primary encoded representation of the current frame is contained. Thus, by having two separate packets available, the decoder can use the error concealment parameter contained in one of these separate packets in case that the packet containing the primary encoded representation of the current frame is lost, corrupted or delayed.

In an embodiment, the decoder may be configured to selectively choose between at least two error concealment modes which use different encoded representations of one or more error concealment parameters for at least partially reconstructing the audio content using the extrapolation-based error concealment. The decoder chooses one of the at least two error concealment modes if the decoder does not get the respective mode, i.e. if the decoder cannot determine or otherwise retrieve the respective mode, from the partial copy (i.e. from the encoded representation of the at least one error concealment parameter). Otherwise, the concealment mode is dictated by the available partial copy, i.e. by the encoded representation of the at least one error concealment parameter. In this case, the encoder already made the choice, while the decoder uses the selected one the at least two modes. In other words, in CA-mode, the encoder decides on the appropriate concealment mode and prepares the partial copy accordingly. If a partial copy is available to the decoder and it should be used for enhancing the concealment, the decoder must stick to the decision made by the encoder, otherwise the info within the partial copy cannot be exploited properly. The decoder only

decides itself on the concealment mode, if no partial copy is available or if the partial copy is not and/or should not be used for other reasons. Accordingly, the decoder provides for a signal specific decoding of one or more error concealment parameters and an enhanced error concealment.

- 5 In an embodiment, at least one of the error concealment modes which uses different encoded representations of one or more error concealment parameters may be a time domain concealment mode wherein the encoded representation of the at least one error concealment parameter comprises at least one of a TCX LTP lag and a classifier information. For example, a first mode which is a time domain concealment mode, could be selected if a time domain
10 signal is present comprising at least a certain characteristic. Otherwise, if the time domain signal does not comprise this certain characteristic, or if the time domain signal comprises a different signal characteristic, a second mode is chosen. Thus, the encoder may provide for a signal specific selection of the error concealment parameters, when the decoder may follow this encoder's selection.
- 15 In an embodiment, at least one of the at least two error concealment modes which uses different encoded representations of one or more error concealment parameters may be a frequency domain concealment mode wherein the encoded representation of the at least one error concealment parameter comprises one or more of an LSF parameter, a TCX global gain and a classifier information. This mode can selectively be chosen by the decoder for providing
20 a good concealment result of an unavailable primary encoded representation containing a frequency domain signal.

In an embodiment, the decoder may use at least a TCX coding scheme. According to this embodiment, the decoder preferably uses TCX decoding for efficiently decoding general audio content, music, background noise, or the like. Thus, the decoder can use TCX specific error
25 concealment parameters for reconstructing a TCX signal in case that the primary encoded representation has been lost, corrupted or delayed.

An embodiment creates an apparatus for error concealment, the apparatus being configured for performing a standard concealment mechanism for a lost frame and to use transmittable parameters to enhance the concealment. Thus, the present invention improves a standard
30 concealment mechanism by using certain parameters.

An embodiment creates an apparatus for error concealment, the apparatus being configured for not having a partial copy that is just a low bitrate version of the primary, but for having a partial copy consisting of multiple key parameters for enhancing the concealment. Thus, bandwidth capacity can be efficiently used.

An embodiment creates an apparatus for error concealment having a receiver comprising a de-jitter buffer for providing a partial redundant copy of a current lost frame if it is available in any of the future frames, wherein the apparatus is configured for reading a partial redundant information bitstream and for updating corresponding parameters. Thus, if a current frame is lost, corrupted or delayed, the inventive apparatus can use the partial redundant copy which has been sent later in time, i.e. with a future frame, in order to reconstruct the frame.

An embodiment creates a switched Coder or decoder, in which there are two or more core coding schemes, whereas for example one uses ACELP for coding speech-like content and the second use TCX for coding general audio content wherein ACELP frames are processed using a partial redundant copy coding and TCX frames are processed using a different approach, wherein in frames that are close to a core coder switch, two special cases can occur, namely: ACELP primary frame with partial copy generated from future TCX frame on top, or TCX primary frame with partial copy generated from future ACELP frame on top, wherein, for these cases, both core coders are configurable to create primary frames in combination with partial copies from the other coder type, without infringing the required total size of a frame, to assure a constant bitrate, or wherein: a first TCX frame after an ACELP frame, where, if this frame gets lost and thus is not available to the decoder, the proposed technique will TCX conceal the frame using partial copy information that has been transported in top of another frame, wherein concealment needs a preceding frame for extrapolating the signal content, ACELP concealment is used (as the previous frame was ACELP) and wherein it is decided already in the encoder, to not put a partial copy on top of a TCX frame after a switch, or where there is a signal-adaptive partial copy selection, where a signal is analyzed before encoding to determine if the usage of partial copy is favorable, wherein if the signal could be concealed satisfyingly well without the help of additional partial copy info within the decoder, but the clean channel performance suffers because of reduced primary frame, a partial copy usage is turned off or a specifically reduced partial copy is used within the encoder. Thus, the inventive coder or decoder is versatile as it provides for a combination of different coding schemes.

An embodiment creates a Transform Domain Coder or decoder, wherein an en-/decoding scheme is used, where at least in some frames redundant coding parameters are embedded in the bitstream and transmitted to the decoder side or wherein a redundant info is delayed by some time and embedded in a packet which is encoded and sent later in time such that the info can be used in the case of the decoder already having the future frame available, and the original frame is lost, corrupted or delayed even more. Thus, by providing redundant coding parameters in the bitstream, a given bandwidth can efficiently be used.

The Transform domain coder or decoder as before may use redundant information comprising ISF/LSF parameters: ISF/LSF parameter representation is used for quantization and coding of

LPC parameters. In TCX the LPC is used to represent the masking threshold. This is an essential parameter and very helpful to have available correctly on decoder side in case of a frame loss. Especially if the ISF/LSFs are coded predictively the concealment quality will improve significantly by having this info available during concealment, because the predictor states on decoder side will stay correct (in sync to encoder) and this will lead to a very quick recovery after the loss; Signal classification: Signal classification is used for signaling the content types: UNVOICED, UNVOICED TRANSITION, VOICED TRANSITION, VOICED and ONSET. Typically this type of classification is used in speech coding and indicating if tonal/predictive components are present in the signal or if the tonal/predictive components are changing. Having this information available on the decoder side during concealment may help to determine the predictability of the signal and thus it can help adjusting the amplitude fade-out speed, the interpolation speed of the LPC parameters; TCX global gain/level: The global gain may be transmitted to easily set the energy of the concealed frame to the correct (encoder determined level) in case it is available; Window information like overlap length; or Spectral peak positions to help tonal concealment.

The terms “redundant”, “redundant copy”, “partial redundant copy” and other combinations of expressions containing the term “redundant” may be used in the sense of providing a “partial” information. A partial information does not contain a redundant, and possibly low-bitrate, representation of a primary-encoded frame, i.e. of an encoded audio signal. Instead, a partial information may contain or comprise parameters, in particular concealment helper parameters which enhance a concealment mechanism that is available at the decoder side, in order to conceal the corresponding primary frame, i.e. the primary-encoded audio data, in case that this primary-encoded frame is defective, e.g. lost, corrupted or delayed. In other words, the terms “redundant” and “partial”, and derivatives thereof, such as e.g. “redundant copy” and “partial copy”, may be used interchangeably within this document, as both terms represent an information that may contain or comprise the aforementioned parameters.

DETAILED DESCRIPTION OF THE EMBODIMENTS

Embodiments of the present invention will be detailed subsequently referring to the appended drawings, in which:

Fig. 1 shows a schematic representation of the inventive encoder,

Fig. 2 shows a schematic representation of an embodiment of an inventive encoder,

Fig. 3 shows a schematic representation of an embodiment of an inventive encoder,

Fig. 4 shows a schematic representation of an embodiment of an inventive encoder,

Fig. 5 shows a schematic representation of an embodiment of an inventive decoder,

Fig. 6 shows a schematic representation of an embodiment showing a concept of partial redundancy in channel aware mode,

Fig. 7 shows a schematic representation of an embodiment showing a concept of partial redundancy in channel aware mode,

5 Fig. 8 shows a schematic representation of an embodiment showing a channel aware encoder framework,

Fig. 9 shows a schematic representation of an embodiment showing a channel aware decoder framework,

Fig. 10 shows a diagram representing Wideband ITU-T P.800 ACR MOS test results, and

10 Fig. 11 shows a diagram representing Super-wideband ITU-T P.800 DCR MOS test results

Figure 1 shows an inventive encoder 1. The encoder 1 is configured to encode audio content 2. In particular, the encoder 1 is configured to encode speech-like content and/or general audio content. The respective encoded audio content 3 is embedded, in at least a frame 4, into a
15 bitstream 5.

The encoder 1 is further configured to embed, at least in some frames 7, parameters 6 in the bitstream 5. These parameters 6 are used to enhance a concealment in case an original frame 4 is lost, corrupted or delayed.

The bitstream 5 is sent to a receiver comprising a decoder.

20 As shown in Figure 2, the encoder 1 is configured to create a primary frame 4b and a partial copy 8b. However, the partial copy 8b is not just a low bitrate version of the primary frame 4b. Instead, the partial copy 8b contains the parameters 6 which enhance the concealment at the decoder side, but, on the other hand, does not include full information for reconstructing an audio content of a defective, e.g. lost, corrupted or delayed primary frame. In other words, the
25 partial copy includes one or more parameters to enhance a decoder-sided error concealment, but not all the information needed for the error concealment.

The encoder 1 is configured to delay the parameters 6 by some time and to embed the parameters 6 in a packet 9 which is encoded and sent later in time than a packet which comprises the primary frame 4b.

30 The encoder 1 may create one or more primary frames 4b, 4c and one or more partial copies 8a, 8b. For example, at least a certain part of the audio content 2 is encoded and embedded into a primary frame 4b. The same part of the audio content 2 is analyzed by the encoder 1 as

to certain signal characteristics. Based thereupon, the encoder 1 determines a selection of the one or more parameters 6 which enhance a concealment on the decoder side. These parameters 6 are embedded in a corresponding "partial copy" 8b.

In other words, the primary frame 4b contains an encoded representation of at least a part of the audio content 2. The corresponding partial copy 8b contains one or more parameters 6 which are used by an error concealment at the decoder side in order to reconstruct the encoded representation of the audio content 2 in case the primary frame 4b is lost, corrupted or delayed.

The primary copy 4b is packed into the transport packet 9 together with a partial copy 8a, wherein the partial copy 8a is the partial copy of an audio content that has been encoded in a primary frame 4a which has already been sent earlier in time. Accordingly, the encoder 1 delayed the parameters 6 by some time. As can be further seen in Figure 2, the partial copy 8b (belonging to primary frame 4b) that follows the partial copy 8a will be packed together with the primary frame 4c in a later transport packet. There may also be one or more further primary frames between the primary frames 4c and 4b.

It is an important feature that the concept described herein uses an en-/decoding scheme where at least in some frames 8a, 8b redundant coding parameters 6 are embedded in the bitstream 5 and transmitted to the decoder side. The redundant info (parameters 6) is delayed by some time and embedded in a packet 9 which is encoded and sent later in time such that the info can be used in the case of the decoder already has the future frame 4b, 8a available, but the original frame 4a is lost, corrupted or delayed even more.

The bitstream 5 may, for example, comprise a constant total bitrate. The encoder 1 may be configured to reduce a primary frame bitrate, i.e. a bitrate that is needed to encode a primary frame 4b, 4c when compared to the constant total bitrate. The bitrate reduction for the primary frames 4b, 4c and a partial redundant frame coding mechanism together determine a bitrate allocation between the primary and redundant frames (partial copies) 4b, 4c, 8a, 8b to be included within the constant total bitrate of the bitstream 5. Thus, the encoder 1 is configured to provide a packet 9 containing a primary frame 4b and a partial copy 8a, wherein the size, i.e. the bitrate of the packet 9 is at or below the constant total bitrate.

In other words, the primary Frame bit-rate reduction and partial redundant frame coding mechanisms together determine the bit-rate allocation between the primary and redundant frames 4b, 4c, 8a, 8b to be included within the constant total bitrate. The overall bit rate of a frame 4b holding partial copy parameters 8a (in addition to primary frames) is not increased.

TCX-coding scheme

According to an embodiment, the encoder 1 is part of a codec using a TCX coding scheme. The inventive encoder 1 uses TCX preferably for coding general audio content. In case of TCX, the partial copy 8a, 8b is used to enhance a frame loss algorithm of an error concealment at the decoder side by transmitting some helper parameters 6.

When using a transform domain codec, embedding redundant info 8a, 8b to TCX frames 4b, 4c may be chosen if:

- The Frame contains a really noisy audio signal. This may be indicated by a low auto correlation measure or by the Frame classifier output being UNVOICED or UNVOICED TRANSITION. UNVOICED or UNVOICED TRANSITION classification indicates a low prediction gain.
- The frame contains a noise floor with sharp spectral lines which are stationary over a longer period of time. This may be detected by a peak detection algorithm which is searching for local maxima in the TCX spectrum (power spectrum or real spectrum) and comparing the result with the result of the peak detection of the previous Frame. In case the peaks did not move it is likely that there are stationary tones which can easily be concealed after having concealed the noise spectrum by post processing the spectrum with a phase extrapolator called tonal concealment.
- In case LTP info is present and the lag is stable over the actual and the past Frame Tonal concealment [6] should be applied at the decoder.

Redundant information (parameters 6) may be:

- ISF/ LSF parameters:

ISF/LSF parameter representation is used for quantization and coding of LPC parameters. In TCX the LPC is used to represent the masking threshold. This is an important parameter and very helpful to have available correctly on decoder side in case of a frame loss. Especially if the ISF/LSFs are coded predictively the concealment quality will improve significantly by having this info available during concealment, because the predictor states on decoder side will stay correct (in sync to encoder) and this will lead to a very quick recovery after the loss.

- Signal classification:

Signal classification is used for signaling the content types: UNVOICED, UNVOICED TRANSITION, VOICED TRANSITION, VOICED and ONSET. Typically this type of classification is used in speech coding and indicating if tonal/ predictive components are present in the signal or if the tonal/predictive components are changing. Having this information available on the decoder side during concealment may help to determine the predictability of the signal and thus it can help adjusting the amplitude fade-out speed, the interpolation speed of the LPC parameters.

- TCX global gain/level:

The global gain may be transmitted to easily set the energy of the concealed frame to the correct (encoder determined) level in case it is available.

- Window information like overlap length.
- Spectral peak positions to help tonal concealment

There is a special case where, at the encoder 1 for frequency domain partial copy, it is checked if the signal 2 contains an onset. If the gain (could be quantized) of the actual frame 4c is more than a certain factor (e.g. 1.6) time the gain of the previous frame 4b and the correlation between the actual frame 4c and the previous frame 4b is low, only a limited (clipped) gain is transmitted. This avoids getting pre echo artefacts in case of concealment. In case of Onset the previous frame 4b is really uncorrelated to the actual frame 4c. Thus, it cannot be relied on the gain computed on the actual frame 4c if concealment is done based on the previous frame 4b spectral bins.

Switched codec scheme (TCX – ACELP)

In a further embodiment, the encoder 1 is part of a switched codec, wherein the switched codec consists of at least two core coding schemes. A first core coding scheme uses ACELP and a second core coding scheme uses TCX. With reference to Figure 3, the encoder 1 comprises a core coder 10 which can switch between ACELP and TCX core coding schemes.

The encoder further comprises an ACELP processor 11 for processing ACELP-coded content 13, and a TCX processor 12 for processing TCX-coded content 14. The ACELP processor 11 is a commonly known processor using a conventional partial copy approach, wherein primary frames 15 are primary coded and redundant frames 16 are redundant-coded. The redundant frames 16 are a low-bitrate version of their corresponding primary frames 15.

The TCX processor 12 processes frames that have been encoded according to the inventive concept. In a first branch 17, the encoded content 3 is provided in the form of primary frames 4b, 4c. In a second branch 18, the parameters 6 which enhance the concealment are provided in the form of "partial copies" 8a, 8b, such as shown in Figure 2. Both the ACELP content 15, 16 and the TCX content 17, 18 are packed into a sequence of transport packets 9, as described before, and sent in the bitstream 5 to the decoder side.

Still with reference to Figure 3, but stated in different words, the usage of the inventive concept is described in combination with a state of the art partial redundant copy based approach in a switched coding system. Such a system consists of two (or more) core coding schemes, whereas one uses ACELP for coding speech-like content and the second use TCX for coding general audio content.

Assuming ACELP frames 15, 16 are processed using traditional partial redundant copy coding and TCX frames 4b, 4c, 8a, 8b are processed using the inventive approach, two main cases will occur, where no special action is needed and the frames 4b, 4c, 8a, 8b, 15, 16 can be processed using the underlying core coder's partial copy approach:

- ACELP primary frame 15 with partial copy 16 generated from future ACELP frame on top
- TCX primary frame 4c with partial copy 8b generated from future TCX frame 4b on top

However, in frames that are close to a core coder switch, two special cases can occur, namely

- ACELP primary frame 15 with partial copy 8 generated from future TCX frame on top
- TCX primary frame 4 with partial copy 16 generated from future ACELP frame on top

For these cases, both core coders need to be configurable to create primary frames 4, 15 in combination with partial copies 8, 16 from the other coder type, without infringing the required total size of a frame, to assure a constant bitrate.

Accordingly, the encoder 1 is configured to create a primary frame 4, 15 of one of the speech-like content type (ACELP) and the general audio content type (TCX) in combination with a partial copy 8, 16 of the other one of the speech-like content type and the general audio content type.

However, there are more specific cases, where a more sophisticated selection of partial copies 8, 16 is appropriate, e.g.:

First TCX frame 4 after an ACELP frame 15:

If this frame 4 gets lost and thus is not available to the decoder, the inventive technique will TCX-conceal the frame 4 using partial copy information (parameters 6) that has been transported in top of another (hopefully not lost) frame. But as concealment needs a preceding frame for extrapolating the signal content, it is preferable in this case to use ACELP concealment (as the previous frame was ACELP) which would make a TCX partial copy unnecessary. Thus it is decided already in the encoder 1, to not put a partial copy 8 on top of a TCX frame 4 after a switch.

Accordingly, the encoder 1 is configured to not put a partial copy 8 on top of a TCX frame 4 after a switch when there is a first TCX frame 4 after an ACELP frame 15.

Signal-adaptive partial copy selection:

The signal (audio content) 2 can be analyzed before encoding to determine if the usage of the inventive partial copy (using parameters 6) is favorable. For example, if the signal 2 could be concealed satisfyingly well without the help of additional partial copy info, i.e. parameters 6, within the decoder, but the clean channel performance suffers because of reduced primary frame 4, the inventive partial copy usage (i.e. embedding parameters 6 in the bitstream 5) can be e.g. turned off or a specifically reduced partial copy 8 can be used within the encoder 1.

Accordingly, the encoder 1 is configured to analyze the signal 2 before encoding and to turn off the partial copy usage or to provide a reduced partial copy based on the analyzed signal 2.

Generally, the encoder 1 is configured to provide partial redundant copies 8 which are constructed in a partial copy mode. In an embodiment, the encoder 1 is configured to choose between multiple partial copy modes which use different amounts of information and/or different parameter sets, wherein the selection of the partial copy mode is based on various parameters.

Construction of partial redundant frame for TCX frame

In case of TCX partial redundant frame type, a partial copy 8 consisting of some helper parameters 6 is used to enhance the frame loss concealment algorithm. In an embodiment, there are three different partial copy modes available, which are RF_TCXFD, RF_TCXTD1 and RF_TCX_TD2. Similar to the PLC mode decision on the decoder side, the selection of the

partial copy mode for TCX is based on various parameters such as the mode of the last two frames, the frame class, LTP pitch and gain. The parameters used for the selection of the mode may be equal to or different from the parameters for enhancing the concealment which are included in the "partial copy".

5

a) Frequency domain concealment (RF_TCXFD) partial redundant frame type

According to an embodiment, at least one of the multiple partial copy modes is a frequency domain ("FD") concealment mode, an example of which is described in the following.

10 29 bits are used for the RF_TCXFD partial copy mode.

- 13 bits are used for the LSF quantizer (e.g. for coding LPC parameters) which is the same as used for regular low rate TCX coding.
- The global TCX gain is quantized using 7 bits.
- The classifier info (e.g. VOICED, UNVOICED, etc.) is coded on 2 bits.

15 b) Time domain concealment (RF_TCXTD1 and RF_TCXTD2) partial redundant frame type

According to an embodiment, at least two of the multiple partial copy modes are different time domain ("TD") concealment modes, an example of which is described in the following. A first time domain concealment mode, namely the partial copy mode RF_TCXTD1 is selected if a frame 4c contains a transient or if the global gain of the frame 4c is (much) lower than the
 20 global gain of the previous frame 4b. Otherwise, the second time domain concealment mode, namely RF_TCXTD2 is chosen.

Overall 18bits of side data are used for both modes.

- 9bits are used to signal the TCX LTP (Long Term Prediction) lag
- 2 bits for signaling the classifier info (e.g. VOICED, UNVOICED, etc.)

25

Time Domain Concealment

Depending on the implementation, the codec could be a transform domain codec only or a switch codec (transform/time domain) using the time domain concealment described in [4] or [5]. Similar to the therein described packet loss concealment mode decision on the decoder

side, the selection of the partial copy mode according to the present invention is based on various parameters, as mentioned above, e.g. the mode of the last two frames, the frame class, LTP pitch and gain.

In the case time domain mode is chosen, the following parameters 6 can be transmitted:

- 5 • In the case LTP data is present, the LTP lag is transmitted,
- a classifier info is signaled (UNVOICED, UNVOICED TRANSITION, VOICED, VOICED TRANSITION, ONSET ...): Signal classification is used for signaling the content types: UNVOICED, UNVOICED TRANSITION, VOICED TRANSITION, VOICED and ONSET. Typically this type of classification is used in speech coding and indicating if
- 10 tonal/predictive components are present in the signal or if the tonal/predictive components are changing. Having this information available on the decoder side during concealment may help to determine the predictability of the signal and thus it can help adjusting the amplitude fade-out speed, the interpolation speed of the LPC parameters and it can control possible usage of high- or low pass filtering of voiced or unvoiced
- 15 excitation signals (e.g. for de-noising).

Optionally, also at least one of the following parameters 6 can be transmitted:

- LPC parameters describing the full spectral range in case of bandwidth extension is used for regular coding,
- LTP Gain,
- 20 • Noise level, and
- Pulse position

Most of the parameters 6 sent, are directly derived from the actual frame 4 coded in the transform domain, so there is no additional complexity caused. But if the complexity is not an issue, then a concealment simulation at the encoder 1 can be added to refine the variable 6

25 that can be sent.

As mentioned above, also multiple modes for the provision of the partial copy 8 can be used. This permits to send different amounts of information or different parameter sets. For example, there are two modes for the time domain (TD). The partial copy mode TD1 could be selected if the frame 4c contains a transient or if the global gain of the frame 4c is much lower than the

30 global gain of the previous frame 4b. Otherwise TD2 is chosen. Then at the decoder, the pitch gain and the code gain will be decreased with two different factors (0.4 and 0.7 accordingly) to avoid having a long stationary signal whenever the original signal 2 was more transient like.

Multiple Frame Loss

There is a further special case, namely the case of multiple frame loss. The pitch decoded from the partial copy 8b shall not be taken into account if the previous frame 4a is lost, because the pitch sent in the bitstream 5 was computed on the encoder side based on the ground truth, but if the previous frame 4a is lost, the synthesis of the previously lost and concealed synthesis might be really different to the encoder ground truth. So it is better in general to not risk relying on the synchronicity of en-/decoder in case of multiple frame loss and fix the pitch to the predicted pitch for the following lost frame instead of using the pitch transmitted.

The inventive concept of the encoder 1 shall be summarized in the following with reference to an embodiment as shown Figure 4.

The encoder 1 receives an input signal which contains audio content 2. The audio content 2 may be speech-like content and/or general audio content such as music, background noise or the like.

The encoder 1 comprises a core coder 10. The core coder 10 can use a core coding scheme for encoding speech-like content, such as ACELP, or a core coding scheme for encoding general audio content, such as TCX. The core coder 10 may also form part of a switched codec, i.e. the core coder 10 can switch between the speech-like content core coding scheme and the general audio content core coding scheme. In particular, the core coder 10 can switch between ACELP and TCX.

As indicated in branch 20, the core coder 10 creates primary frames 4 which comprise an encoded representation of the audio content 2.

The encoder 1 may further comprise a partial redundant frame provider 21. As indicated in branch 30, the core coder 10 may provide one or more parameters 6 to the partial redundant frame provider 21. These parameters 6 are parameters which enhance a concealment at the decoder side.

Additionally or alternatively, the encoder 1 may comprise a concealment parameter extraction unit 22. The concealment parameter extraction unit 22 extracts the concealment parameters 6 directly from the audio signal, i.e. from the content 2, as indicated in branch 40. The concealment parameter extraction unit 22 provides the extracted parameters 6 to the partial redundant frame provider 21.

The encoder 1 further comprises a mode selector 23. The mode selector 23 selectively chooses a concealment mode, which is also called partial redundant copy mode. Depending

on the partial redundant copy mode, the mode selector 23 determines which parameters 6 are suitable for an error concealment at the decoder side.

Therefore, the core coder 10 analyzes the signal, i.e. the audio content 2 and determines, based on the analyzed signal characteristics, certain parameters 24 which are provided to the mode selector 23. These parameters 24 are also referred to as mode selection parameters 24. For example, mode selection parameters can be at least one of a frame class, the mode of the last two frames, LTP pitch and LTP gain. The core coder 10 provides these mode selection parameters 24 to the mode selector 23.

Based on the mode selection parameters 24, the mode selector 23 selects a partial redundant copy mode. The mode selector 23 may selectively choose between three different partial redundant copy modes. In particular, the mode selector 23 may selectively choose between a frequency domain partial redundant copy mode and two different time domain partial redundant copy modes, e.g. TD1 and TD2, for example as described above.

As indicated in branch 50, the mode selection information 25, i.e. the information regarding the selected partial redundant copy mode, is provided to the partial redundant frame provider 21. Based on the mode selection information 25, the partial redundant frame provider 21 selectively chooses parameters 6 that will be used, at the decoder side, for error concealment. Therefore, the partial redundant frame provider 21 creates and provides partial redundant frames 8 which contain an encoded representation of said error concealment parameters 6.

Stated differently, the partial redundant frame provider 21 provides signal specific partial redundant copies. These partial redundant copies are provided in partial redundant frames 8, wherein each partial redundant frame 8 contains at least one error concealment parameter 6.

As indicated at the branches 20 and 60, the encoder 1 combines the primary frames 4 and the partial redundant frames 8 into the outgoing bitstream 5. In the case of a packet-based network, primary frames 4 and partial redundant frames 8 are packed together into a transport packet, which is sent in the bitstream to the decoder side. However, it is to be noted that the primary frame 4c of a current audio frame is packed into a packet 9 together with a partial redundant frame 8b (containing only the parameters 6 for enhancing a concealment) of a previous frame (i.e. a frame that has already been sent earlier in time).

The bitstream 5 comprises a constant total bitrate. In order to ensure that the bitstream 5 is at or below the constant total bitrate, the encoder 1 controls the bitrate of the transport packet containing the combination of the primary frame and the partial redundant frame 8. Additionally or alternatively, the encoder 1 may comprise a bitrate controller 26 that takes over this functionality.

In other words, the encoder 1 is configured to combine an encoded representation 8 of the at least one concealment parameter 6 of a current frame with a primary encoded representation 4 of a future frame (i.e. a frame that will be sent later in time than the current frame). Thus, the encoded representation 8 of the at least one error concealment parameter 6 of a current frame is sent with a time delay relative to the primary encoded representation 4 of this current frame.

Stated differently, and still with reference to Figure 4, in a first step, content 2a is encoded and provided as a primary frame 4a. Its corresponding one or more error concealment parameters 6a are selected and provided as a partial redundant frame 8a. Then, in a second step, a subsequent content 2b is encoded and provided as a (subsequent) primary frame 4b and its one or more error concealment parameters 6b are selected and provided as a (subsequent) partial redundant frame 8b. Now, the encoder 1 combines the partial redundant frame 8a (of the current content) with the primary frame 4b (of the subsequent content) into a common transport packet 9b. Accordingly, if the preceding packet 9a containing primary frame 4a is lost, corrupted or delayed, its partial redundant frame 8a, which is sent later in time within the above mentioned subsequent transport packet 9b (containing partial redundant frame 8a and primary frame 4b), can be used at the decoder side for concealment of the audio content that was originally contained in an encoded representation in (defective) primary frame 4a.

DESCRIPTION OF THE DECODER

According to an embodiment, the invention uses packet-switched, or packet-based networks. In this case, frames are sent in transport packets 9a, 9b, as shown in Figure 5. Transport packet 9a contains a primary frame 4b and a partial copy 8a. Transport packet 9b contains a primary frame 4c and a partial copy 8b.

Stated differently, a partial copy 8a is an encoded representation of at least one error concealment parameter 6 of a current frame. The at least one error concealment parameter 6 has been selectively chosen by the encoder 1, as described before with reference to Figures 1 to 4. The at least one error concealment parameter 6 enhances a concealment at the decoder 31, as will be described in more detail below.

At the decoder 31, there may be two different cases regarding the transmitted frames 4, 8 or transport packets 9a, 9b, respectively.

Standard decoding of primary encoded representations

In a first case, indicated by branch 70, the transmitted transport packets 9a, 9b are received in the correct order, i.e. in the same order as they have been sent at the encoder side.

5 The decoder 31 comprises a decoding unit 34 for decoding the transmitted encoded audio content 2 contained in the frames. In particular, the decoding unit 34 is configured to decode the transmitted primary encoded representations 4b, 4c of certain frames. Depending on the encoding scheme of the respective frame, the decoder 31 may use the same scheme for decoding, i.e. a TCX decoding scheme for general audio content or an ACELP decoding scheme for speech-like content. Thus, the decoder 31 outputs a respectively decoded audio
10 content 35.

Enhanced error concealment using encoded representations of at least one error
concealment parameter

15 A second case may occur if a primary encoded representation 4 of a frame is defective, i.e. if a primary encoded representation 4 is lost, corrupted or delayed (for example because the transport packet 9a is lost, corrupted or delayed longer than a buffer length of the decoder), such as indicated by branch 80. The audio content will then have to be at least partly reconstructed by error concealment.

20 Therefore, the decoder 31 comprises a concealment unit 36. The concealment unit 36 may use a concealment mechanism which is based on a conventional concealment mechanism, wherein, however, the concealment is enhanced (or supported) by one or more error concealment parameters 6 received from the encoder 1. According to an embodiment of the invention, the concealment unit 36 uses an extrapolation-based concealment mechanism, such as described in patent applications [4] and [5], which are incorporated herein by
25 reference.

Said extrapolation-based error concealment mechanism is used in order to reconstruct audio content that was available in a primary encoded representation 4 of a frame, in the case that this primary encoded representation 4 is defective, i.e. lost, corrupted or delayed. The inventive concept uses the at least one error concealment parameter 6 to enhance these conventional
30 error concealment mechanisms.

This shall be explained in more detail with reference to the embodiment shown in Figure 5. The decoder 31 normally receives a transport packet 9a and a transport packet 9b. Transport

packet 9a contains a primary encoded representation 4b of a current frame and an encoded representation 8a of at least one error concealment parameter 6 of a preceding frame (not shown). Transport packet 9b contains an encoded representation 8b of at least one error concealment parameter 6 of the current frame for enhancing a decoder-sided extrapolation-based error concealment of the current frame. Transport packet 9b further contains a primary encoded representation 4c of a subsequent frame, i.e. a frame following (directly or with one or more frames in between) the current frame.

Stated differently, the encoded representation 8b of the at least one error concealment parameter 6 for reconstructing the defective audio content of the current frame is contained in transport packet 9b, while the primary encoded representation 4b of this current frame is contained in transport packet 9a.

If it is detected by the decoder 31 that, for example, the primary encoded representation 4b of the current frame is defective, i.e. lost, corrupted or delayed, the defective audio content is reconstructed by using the afore-mentioned available error concealment mechanism.

According to the present invention, the available error concealment mechanism is enhanced by using the at least one error concealment parameter 6 during error concealment.

For this reason, the decoder 31 extracts the at least one error concealment parameter 6 from the encoded representation 8b contained in transport packet 9b. Based on the at least one parameter 6 that has been extracted, the decoder 31 selectively chooses between at least two concealment modes for at least partially reconstructing the defective audio content (in the sense that a concealed audio content is provided which is expected to be somewhat similar to the audio content of the lost primary encoded representation). In particular, the decoder 31 can choose between a frequency domain concealment mode and at least one time domain concealment mode.

Frequency domain concealment (RF TCXFD) partial redundant frame type

In case of a frequency domain concealment mode, the encoded representation 8b of the at least one error concealment parameter 6 comprises one or more of an ISF/LSF parameter, a TCX global gain, a TCX global level, a signal classifier information, a window information like overlap length and spectral peak positions to help tonal concealment.

The respective extracted one or more parameters 6 are fed to the error concealment unit 36 which uses the at least one parameter 6 for enhancing the extrapolation-based error concealment in order to at least partially reconstruct the defective audio content. As a result, the decoder 31 outputs the concealed audio content 35.

An embodiment of the present invention, which uses an example of a frequency domain concealment, is described below, wherein

29 bits are used for the RF_TCXFD partial copy mode (i.e. 29 bits are included in the encoded representation of error concealment parameters 6 and are used by the concealment unit 36).

- 5 • 13 bits are used for the LSF quantizer which is the same as used for regular low rate TCX coding.
- The global TCX gain is quantized using 7 bits.
- The classifier info is coded on 2 bits.

Time domain concealment (RF_TCXTD1 and RF_TCXTD2) partial redundant frame type

- 10 In case of a time domain concealment mode, the decoder 31 may selectively choose between at least two different time domain concealment modes in order to at least partially reconstruct the defective audio content.

For example, a first mode RF_TCXTD1 is selected if the frame contains a transient or if the global gain of the frame is much lower than the global gain of the previous frame. Otherwise,

- 15 a second mode RF_TCXTD2 is chosen.

In case of a time domain concealment mode, the encoded representation 8b of the at least one error concealment parameter 6 comprises one or more of an LSF parameter, a TCX LTP lag, a classifier information, LPC parameters, LTP gain, Noise Level and Pulse Position. The respective extracted one or more parameters 6 are fed to the error concealment unit 36 which

20 uses the at least one parameter 6 for enhancing the extrapolation-based error concealment in order to at least partially reconstruct (or approximate) the defective audio content. As a result, the decoder 31 outputs the concealed audio content 35.

An embodiment of the present invention, which uses an example of a time domain concealment, is described below, wherein

- 25 Overall 18bits of side data (i.e. of parameters 6) are used for both modes.

- 9bits are used to signal the TCX LTP lag
- 2 bits for signaling the classifier info

The decoder 31 may be part of a codec using a TCX decoding scheme for decoding and/or concealing TCX frames, as described above. The decoder 31 may also be part of a codec

using an ACELP coding scheme for decoding and/or concealing ACELP frames. In case of ACELP coding scheme, the encoded representation 8b of the at least one error concealment parameter 6 may comprise one or more of adaptive codebook parameters and a fixed codebook parameter.

5 According to the invention, in the decoder 31 the type of the encoded representation of the at least one error concealment parameter 6 of a current frame 4b is identified and decoding and error concealment is performed based on whether only one or more adaptive codebook parameters (e.g. ACELP), only one or more fixed codebook parameters (e.g. ACELP), or one or more adaptive codebook parameters and one or more fixed codebook parameters, TCX
10 error concealment parameters 6, or Noise Excited Linear Prediction parameters are coded. If the current frame 4b or a previous frame 4a is concealed by using an encoded representation of at least one error concealment parameter 6 of the respective frame, the at least one error concealment parameter 6 of the current frame 4b, such as LSP parameters, the gain of adaptive codebook, fix codebook or the BWE gain, is firstly obtained and then processed in
15 combination with decoding parameters, classification information or spectral tilt from previous frames of the current frame 4b, or from future frames of the current frame 4b, in order to reconstruct the output signal 35, as described above. Finally, the frame is reconstructed based on the concealment scheme (e.g. time-domain concealment or frequency-domain concealment). The TCX partial info is decoded, but in contrast to an ACELP partial copy mode,
20 the decoder 31 is run in concealment mode. The difference to the above described conventional extrapolation-based concealment is that the at least one error concealment parameter 6 which is available from the bitstream 5 is directly used and not derived by said conventional concealment.

25

First EVS-embodiment

The following description passages provide a summary of the inventive concept with respect to the synergistic interaction between encoder 1 and decoder 31 using a so-called EVS (Enhanced Voice Services) Codec.

30

Introduction to EVS-embodiment

EVS (Enhanced Voice Services) offers partial redundancy based error robust channel aware mode at 13.2 kbps for both wideband and super-wideband audio bandwidths. Depending on the criticality of the frame, the partial redundancy is dynamically enabled or disabled for a particular frame, while keeping a fixed bit budget of 13.2 kbps.

Principles of Channel Aware Coding

In a VoIP system, packets arrive at the decoder with random jitters in their arrival time. Packets may also arrive out of order at the decoder. Since the decoder expects to be fed a speech packet every 20 msec to output speech samples in periodic blocks, a de-jitter buffer [6] is required to absorb the jitter in the packet arrival time. Larger the size of the de-jitter buffer, the better is its ability to absorb the jitter in the arrival time and consequently, fewer late arriving packets are discarded. Voice communications is also a delay critical system and therefore it becomes essential to keep the end to end delay as low as possible so that a two way conversation can be sustained.

The design of an adaptive de-jitter buffer reflects the above mentioned trade-offs. While attempting to minimize packet losses, the jitter buffer management algorithm in the decoder also keeps track of the delay in packet delivery as a result of the buffering. The jitter buffer management algorithm suitably adjusts the depth of the de-jitter buffer in order to achieve the trade-off between delay and late losses.

With reference to Figure 6, EVS channel aware mode uses partial redundant copies 8a of current frames 4a along with a future frame 4b for error concealment. The partial redundancy technology transmits partial copies 8a of the current frame 4a along with a future frame 4b with the hope that in the event of the loss of the current frame 4a (either due to network loss or late arrival) the partial copy 8a from the future frame 4b can be retrieved from the jitter buffer to improve the recovery from the loss.

The difference in time units between the transmit time of the primary copy 4a of a frame and the transmit time of the redundant copy 8a of the frame (piggy backed onto a future frame 4b) is called the FEC offset. If the depth of the jitter buffer at any given time is at least equal to the FEC offset, then it is quite likely that the future frame is available in the de-jitter buffer at the current time instance. The FEC offset is a configurable parameter at the encoder which can be dynamically adjusted depending on the network conditions.

The concept of partial redundancy in EVS with FEC offset equal to [7] is shown in Figure 6.

The redundant copy 8a is only a partial copy that includes just a subset of parameters that are most critical for decoding or arresting error propagation.

The EVS channel aware mode transmits redundancy in-band as part of the codec payload as opposed to transmitting redundancy at the transport layer (e.g., by including multiple packets

in a single RTP payload). Including the redundancy in-band allows the transmission of redundancy to be either channel controlled (e.g., to combat network congestion) or source controlled. In the latter case, the encoder can use properties of the input source signal to determine which frames are most critical for high quality reconstruction at the decoder and selectively transmit redundancy for those frames only. Another advantage of in-band redundancy is that source control can be used to determine which frames of input can best be coded at a reduced frame rate in order to accommodate the attachment of redundancy without altering the total packet size. In this way, the channel aware mode includes redundancy in a constant-bit-rate channel (13.2 kbps).

Bit-Rate Allocation for Primary and Partial Redundant Frame Coding

Primary frame bit-rate reduction

A measure of compressibility of the primary frame is used to determine which frames can best be coded at a reduced frame rate. For TCX frame the 9.6 kbps setup is applied for WB as well as for SWB. For ACELP the following apply. The coding mode decision coming from the signal classification algorithm is first checked. Speech frames classified for Unvoiced Coding (UC) or Voiced Coding (VC) are suitable for compression. For Generic Coding (GC) mode, the correlation (at pitch lag) between adjacent sub-frames within the frame is used to determine compressibility. Primary frame coding of upper band signal (i.e., from 6.4 to 14.4 kHz in SWB and 6.4 to 8 kHz in WB) in channel aware mode uses time-domain bandwidth extension (TBE). For SWB TBE in channel aware mode, a scaled down version of the non-channel aware mode framework is used to obtain a reduction of bits used for the primary frame. The LSF quantization is performed using an 8-bit vector quantization in channel aware mode while a 21-bit scalar quantization based approach is used in non-channel aware mode. The SWB TBE primary frame gain parameters in channel aware mode are encoded similar to that of non-channel aware mode at 13.2 kbps, i.e., 8 bits for gain parameters. The WB TBE in channel aware mode uses similar encoding as used in 9.6 kbps WB TBE of non-channel aware mode, i.e., 2 bits for LSF and 4 bits for gain parameters.

Partial Redundant Frame Coding

The size of the partial redundant frame is variable and depends on the characteristics of the input signal. Also criticality measure is an important metric. A frame is considered as critical to protect when loss of the frame would cause significant impact to the speech quality at the

receiver. The criticality also depends on if the previous frames were lost or not. For example, a frame may go from being non-critical to critical if the previous frames were also lost. Parameters computed from the primary copy coding such as coder type classification information, subframe pitch lag, factor M etc are used to measure the criticality of a frame. The threshold, to determine whether a particular frame is critical or not, is a configurable parameter at the encoder which can be dynamically adjusted depending on the network conditions. For example, under high FER conditions it may be desirable to adjust the threshold to classify more frames as critical. Partial frame coding of upper band signal relies on coarse encoding of gain parameters and interpolation/extrapolation of LSF parameters from primary frame. The TBE gain parameters estimated during the primary frame encoding of the (n – FEC offset)-th frame is re-transmitted during the n-th frame as partial copy information. Depending on the partial frame coding mode, i.e., GENERIC or VOICED or UNVOICED, the re-transmission of the gain frame, uses different quantization resolution and gain smoothing.

The following sections describe the different partial redundant frame types and their composition.

Construction of partial redundant frame for Generic and Voiced Coding modes

In the coding of the redundant version of the frame, a factor M is determined based on the adaptive and fixed codebook energy.

$$M = \frac{(E(ACB) + E(FCB)/E(ACB) - E(FCB)) + 1}{4}$$

In this equation, $E(ACB)$ denotes the adaptive codebook energy and $E(FCB)$ denotes the fixed codebook energy. A low value of M indicates that most of the information in the current frame is carried by the fixed codebook contribution. In such cases, the partial redundant copy (RF_NOPRED) is constructed using one or more fixed codebook parameters only (FCB pulses and gain). A high value of M indicates that most of the information in the current frame is carried by the adaptive codebook contribution. In such cases, the partial redundant copy (RF_ALLPRED) is constructed using one or more adaptive codebook parameters only (pitch lag and gain). If M takes mid values then a mixed coding mode is selected where one or more adaptive codebook parameters and one or more fixed codebook parameters are coded (RF_GENPRED). Under Generic and Voiced Coding modes, the TBE gain frame values are typically low and demonstrate less variance. Hence a coarse TBE gain frame quantization with gain smoothing is used.

Construction of partial redundant frame for Unvoiced Coding mode

The low bit-rate Noise Excited Linear Prediction coding scheme is used to construct a partial redundant copy for an unvoiced frame type (RF_NELP). In Unvoiced coding mode, the TBE gain frame has a wider dynamic range. To preserve this dynamic range, the TBE gain frame quantization in Unvoiced coding mode uses a similar quantization range as that of the one used in the primary frame.

Construction of partial redundant frame for TCX frame

In case of TCX partial redundant frame type, a partial copy consisting of some helper parameters is used to enhance the frame loss concealment algorithm. There are three different partial copy modes available, which are RF_TCXFD, RF_TCXTD1 and RF_TCX_TD2. Similar to the PLC mode decision on the decoder side, the selection of the partial copy mode for TCX is based on various parameters such as the mode of the last two frames, the frame class, LTP pitch and gain.

Frequency domain concealment (RF_TCXFD) partial redundant frame type

29 bits are used for the RF_TCXFD partial copy mode.

- 13bits are used for the LSF quantizer which is the same as used for regular low rate TCX coding.
- The global TCX gain is quantized using 7 bits.
- The classifier info is coded on 2 bits.

Time domain concealment (RF_TCXTD1 and RF_TCXTD2) partial redundant frame type

The partial copy mode RF_TCXTD1 is selected if the frame contains a transient or if the global gain of the frame is much lower than the global gain of the previous frame. Otherwise RF_TCXTD2 is chosen.

Overall 18bits of side data are used for both modes.

- 9bits are used to signal the TCX LTP lag
- 2 bits for signalling the classifier info

RF_NO_DATA partial redundant frame type

This is used to signal a configuration where the partial redundant copy is not sent and all bits are used towards primary frame coding.

The primary frame bit-rate reduction and partial redundant frame coding mechanisms together
5 determine the bit-rate allocation between the primary and redundant frames to be included within a 13.2 kbps payload.

Decoding

At the receiver, the de-jitter buffer provides a partial redundant copy of the current lost frame
10 if it is available in any of the future frames. If present, the partial redundant information is used to synthesize the lost frame. In the decoding, the partial redundant frame type is identified and decoding performed based on whether only one or more adaptive codebook parameters, only one or more fixed codebook parameters, or one or more adaptive codebook parameters and one or more fixed codebook parameters, TCX frame loss concealment helper parameters, or
15 Noise Excited Linear Prediction parameters are coded. If current frame or previous frame is a partial redundant frame, the decoding parameter of current frame such as LSP parameters, the gain of adaptive codebook, fix codebook or the BWE gain, is firstly obtained and then post-processed according to decoding parameters, classification information or spectral tilt from previous frames of current frame, or future frames of current frame. The post-processed
20 parameters are used to reconstruct the output signal. Finally, the frame is reconstructed based on the coding scheme. The TCX partial info is decoded, but in contrast to ACELP partial copy mode, the decoder is run in concealment mode. The difference to regular concealment is just that the parameters available from the bitstream are directly used and not derived by concealment.

Channel aware mode encoder configurable parameters

The channel aware mode encoder may use the following configurable parameters to adapt its operation to track the channel characteristics seen at the receiver. These parameters maybe
30 computed at the receiver and communicated to the encoder via a receiver triggered feedback mechanism.

Optimal partial redundancy offset (ρ): The difference in time units between the transmit time of the primary copy of a frame (n) and the transmit time of the redundant copy of that frame

which is piggy backed onto a future frame (n+X) is called the FEC offset X. The optimal FEC offset is a value which maximizes the probability of availability of a partial redundant copy when there is a frame loss at the receiver.

Frame erasure rate indicator (p) having the following values: *LO*(low) for FER rates <5% or *HI* (high) for FER>5%. This parameter controls the threshold used to determine whether a particular frame is critical or not. Such an adjustment of the criticality threshold is used to control the frequency of partial copy transmission. The *HI* setting adjusts the criticality threshold to classify more frames as critical to transmit as compared to the *LO* setting.

It is noted that these encoder configurable parameters are optional with default set to p=*HI* and $\rho=3$.

Second EVS-embodiment

The following description passages describe an exemplary embodiment of the inventive concept which is used in packet-switched networks, such as Voice-over-IP (VoIP), Voice-over-LTE (VoLTE) or Voice-over-WiFi (VoWiFi).

A highly error resilient mode of the newly standardized 3GPP EVS speech codec is described. Compared to the AMR-WB codec and other conversational codecs, the EVS channel aware mode offers significantly improved error resilience in voice communication over packet-switched networks such as Voice-over-IP (VoIP) and Voice-over-LTE (VoLTE). The error resilience is achieved using a form of in-band forward error correction. Source-controlled coding techniques are used to identify candidate speech frames for bitrate reduction, leaving spare bits for transmission of partial copies of prior frames such that a constant bit rate is maintained. The self-contained partial copies are used to improve the error robustness in case the original primary frame is lost or discarded due to late arrival. Subjective evaluation results from ITU-T P.800 Mean Opinion Score (MOS) tests are provided, showing improved quality under channel impairments as well as negligible impact to clean channel performance.

Introduction

In packet-switched networks, packets may be subjected to varying scheduling and routing conditions, which results in time-varying end-to-end delay. The delay jitter is not amenable to most conventional speech decoders and voice post-processing algorithms that typically expect the packets to be received at fixed time intervals. Consequently, a de-jitter buffer (also referred

to as Jitter Buffer Management (JBM) [8], [13]) is typically used in the receiving terminal to remove jitter and deliver packets to the decoder in the correct sequential order.

The longer the de-jitter buffer, the better its ability to remove jitter and the greater the likelihood that jitter can be tolerated without discarding packets due to late arrival (or, buffer underflow).

5 However, end-to-end delay is a key determiner of call quality in conversational voice networks, and the ability of the JBM to absorb jitter without adding excessive buffering delay is an important requirement. Thus, a trade-off exists between JBM delay and the jitter induced packet loss at the receiver. JBM designs have evolved to offer increasing levels of performance while maintaining minimal average delay [8]. Aside from delay jitter, the other primary
10 characteristic of packet-switched networks is the presence of multiple consecutive packet losses (error bursts), which are more commonly seen than on circuit switched networks. Such bursts can result from bundling of packets at different network layers, scheduler behavior, poor radio frequency coverage, or even a slow-adapting JBM. However, the de-jitter buffer—an essential component for VoIP—can be leveraged for improved underflow prevention and more
15 sophisticated packet loss concealment [8]. One such technique is to use forward error correction by transmitting encoded information redundantly for use when the original information is lost at the receiver.

Channel aware mode in the EVS codec

20 The EVS Channel Aware mode introduces a novel technique for transmitting redundancy in-band as part of the codec payload in a constant bitrate stream, and is implemented for wideband (WB) and super-wideband (SWB) at 13.2 kbps. This technique is in contrast to prior codecs, for which redundancy is typically added as an afterthought by defining mechanisms to transmit redundancy at the transport layer. For example, the AMR-WB RTP payload format
25 allows for bundling of multiple speech frames to include redundancy into a single RTP payload [9]. Alternatively, RTP packets containing single speech frames can be simply re-transmitted at a later time.

Figure 7 depicts the concept of partial redundancy in the EVS channel aware mode. The idea is to encode and transmit the partial redundant copy 8a associated with the N-th frame, along
30 with the primary encoding 4b of the (N+K)-th frame. The offset parameter, K, which determines the separation between the primary 4 and partial frames 8 is also transmitted along with the partial copy 8. In the packet-switched network, if the N-th frame 4a packet is lost, then the de-jitter buffer 71 is inspected for the availability of future packets. If available, then the transmitted offset parameter is used to identify the appropriate future packet for partial copy extraction and

synthesis of the lost frame. An offset of 3 is used as an example to show the process in Figure 7. The offset parameter can be a fixed value or can be configured at the encoder based on the network conditions. Including the redundancy in-band in EVS Channel Aware mode allows the transmission of redundancy to be either channel-controlled (e.g., to combat network congestion) or source-controlled. In the latter case, the encoder can use properties of the input source signal to determine the frames that are most critical for high quality reconstruction and selectively transmit redundancy for those frames only. Furthermore, the encoder can also identify the frames that can be best coded at a reduced bitrate in order to accommodate the attachment of redundancy while keeping the bit-stream at a constant 13.2 kbps rate. These new techniques significantly improve the performance under degraded channel conditions while maintaining the clean channel quality.

Channel aware encoding

Figure 8 shows a high level description of the channel aware encoder 1. The input audio 2 that is sampled at either 16 kHz (WB) or 32 kHz (SWB) is segmented into frames of 20 msec. A “pre-processing” stage 81 is used to resample the input frame to 12.8 kHz and perform steps such as voice activity detection (VAD) and signal classification [16]. Based on certain analysis parameters (e.g., normalized correlation, VAD, frame type, and pitch lag), the “Redundant frame (RF) configuration” module 82 determines:

1. the compressibility of the current frame 4b, i.e., if the current frame 4b can allow for bitrate reduction, with minimal perceptual impact, to enable the inclusion of a partial copy 8a associated with a previous frame 4a, and
2. the RF frame type classification which controls the number of bits needed to faithfully reconstruct the current frame 4b through the partial copy 8b that is transmitted in a future frame 4c. In Figure 8, the partial copy 8b is transmitted along with a future primary copy 4c at a frame erasure concealment (FEC) offset of 2 frames.

Strongly-voiced and unvoiced frames are suitable for carrying partial copies of a previous frame with negligible perceptual impact to the primary frame quality. If the current frame is allowed to carry the partial copy, it is signaled by setting *RfFlag* in the bit stream to 1, or 0 otherwise. If the *RfFlag* is set to 1, then the number of bits, B_{primary} , available to encode the current primary frame is determined by compensating for the number of bits, BRF, already used up by the accompanying partial copy, i.e., $B_{\text{primary}} = 264 - \text{BRF}$ at 13.2 kbps constant total

bit rate. The number of bits, BRP, can range from 5 to 72 bits depending on frame criticality and RF frame type (Section 3.2).

Primary frame coding

The “primary frame coding” module 83 shown in Figure 8, uses the ACELP coding technology [21], [23] to encode the low band core up to 6.4 kHz while the upper band that is beyond 6.4 kHz and up to the Nyquist frequency is encoded using the Time-domain Bandwidth Extension (TBE) technology [17]. The upper band is parameterized into LSPs and gain parameters to capture both the temporal evolution per sub-frame as well as over an entire frame [17]. The “primary frame coding” module 83 also uses the MDCT-based Transform Coded Excitation (TCX) and Intelligent Gap Filling (IGF) coding technologies [11], [18] to encode the background noise frames and mixed/music content more efficiently. An SNR-based open-loop classifier [22] is used to decide whether to choose the ACELP/TBE technology or the TCX/IGF technology to encode the primary frame.

Dietz et al., [16] give an overview of various advancements to the EVS primary modes that further improve the coding efficiency of the ACELP technology beyond the 3GPP AMR-WB coding efficiency [21]. The EVS Channel Aware mode leverages these ACELP and TCX core advancements for primary frame encoding. Additionally, as the partial copy uses varying number of bits across frames, the primary frame encoding also needs to correspondingly accommodate for an adaptive bit allocation.

Redundant frame coding

The “redundant frame (RF) coding” module 84 performs compact re-encoding of only those parameters that are critical to protect. The set of critical parameters are identified based on the frame’s signal characteristics and are re-encoded at a much lower bitrate (e.g., less than 3.6 kbps). The “bit packer” module 85 arranges the primary frame bit-stream 86 and the partial copy 87 along with certain RF parameters such as RF frame type and FEC offset (see Table I) at fixed locations in the bit-stream.

TABLE I
BIT ALLOCATION FOR CHANNEL AWARE CODING AT 13.2 KBPS

Core coder		ACELP		TCX-IGF
Bandwidth		WB	SWB	
Signalling information (<i>bwidth, coder type, Rflag</i>)		5		
Primary frame	Core	181-248	169-236	232-254
	TBE	6	18	
Partial frame	Core	0-62	0-62	0-22
	TBE	0-5	0-5	
	FEC offset	2		
	RF frame type	3		

A frame is considered as critical to protect when loss of that frame would cause significant impact to the speech quality at the receiver. The threshold to determine whether a particular frame is critical or not is a configurable parameter at the encoder, which can be dynamically adjusted depending on the network conditions. For example, under high FER conditions it may be desirable to adjust the threshold to classify more frames as critical. The criticality may also depend on the ability to quickly recover from the loss of a previous frame. For example if the current frame depends heavily on the previous frame's synthesis, then the current frame may get re-classified from being non-critical to critical in order to arrest the error propagation in case the previous frame were to be lost at the decoder.

a) ACELP Partial Frame Encoding

For ACELP frames, the partial copy encoding uses one of the four RF frame types, RF_NOPRED, RF_ALLPRED, RF_GENPRED, and RF_NELP depending on the frame's signal characteristics. Parameters computed from the primary frame coding such as frame type, pitch lag, and factor τ are used to determine the RF frame type and criticality, where

$$\tau = 0,25 \left(\frac{E_{ACB} - E_{FCB}}{E_{ACB} + E_{FCB}} + 1 \right)$$

E_{ACB} denotes the adaptive codebook (ACB) energy, and E_{FCB} denotes the fixed codebook (FCB) energy. A low value of τ (e.g., 0.15 and below) indicates that most of the information in the current frame is carried by the FCB contribution. In such cases, the RF_NOPRED partial copy encoding uses one or more FCB parameters (e.g., FCB pulses and gain) only. On the other hand, a high value of τ (e.g., 0.35 and above) indicates that most of the information in the current frame is carried by the ACB contribution. In such cases, the RF_ALLPRED partial copy encoding uses one or more ACB parameters (e.g., pitch lag and gain) only. If τ is in the range of [0.15, 0.35], then a mixed coding mode RF_GENPRED uses both ACB and FCB parameters for partial copy encoding. For the UNVOICED frames, low bitrate noise-excited linear prediction (NELP) [16] is used to encode the RF_NELP partial copy. The upper band partial copy coding relies on coarse encoding of gain parameters and extrapolation of LSF parameters from the previous frame [11].

b) TCX Partial Frame Encoding

In order to get a useful TCX partial copy, many bits would have to be spent for coding the MDCT spectral data, which reduces the available number of bits for the primary frame significantly and thus degrades the clean channel quality. For this reason, the number of bits

for TCX primary frames is kept as large as possible, while the partial copy carries a set of control parameters, enabling a highly guided TCX concealment.

The TCX partial copy encoding uses one of the three RF frame types, RF_TCXFD, RF_TCXTD1, and RF_TCXTD2. While the RF_TCXFD carries control parameters for enhancing the frequency-domain concealment, the RF_TCXTD1 and RF_TCXTD2 are used in time-domain concealment [20]. The TCX RF frame type selection is based on the current and previous frame's signal characteristics, including pitch stability, LTP gain and the temporal trend of the signal. Certain critical parameters such as the signal classification, the LSPs, the TCX gain and pitch lag are encoded in the TCX partial copy.

In background noise or in inactive speech frames, a non-guided frame erasure concealment is sufficient to minimize the perceptual artifacts due to lost frames. An RF_NO_DATA is signaled indicating the absence of a partial copy in the bit-stream during the background noise. In addition, the first TCX frame after a switch from ACELP frame, also uses an RF_NODATA due to lack of extrapolation data in such a coding type switching scenario.

Channel aware decoding

Figure 9 represents a high level depiction of the channel aware decoder 31. At the receiver 90, if the current frame 91 is not lost, the JBM 95 provides the packet for "primary frame decoding" 96 and disregards any RF (Redundant Frame) information present in the packet. In case the current frame is lost, and a future frame 94 is available in the de-jitter buffer, then the JBM 95 provides the packet for "partial frame decoding" 97. If a future frame 93 is not available in the de-jitter buffer, then a non-guided erasure concealment [20] is performed.

Interface with JBM

As described earlier, if the N -th frame is not available (lost or delayed) at the play-out time, the JBM is checked for the availability of a future $(N + K)$ -th frame that contains the partial redundancy of the current frame where $K \in \{2, 3, 5, 7\}$. The partial copy of a frame typically arrives after the primary frame. JBM delay adaptation mechanisms are used to increase the likelihood of availability of partial copies in the future frames, especially for higher FEC offsets of 5 and 7. The EVS JBM conforms to the delay-jitter requirements specified by the 3GPP TS 26.114 [10] for all the EVS modes including the channel aware mode.

In addition to the above described functionality, the EVS JBM [13] computes the channel error rate and an optimum FEC offset, K , that maximizes the availability of the partial redundant copy based on the channel statistics. The computed optimum FEC offset and the channel error rate can be transmitted back to the encoder through a receiver feedback mechanism (e.g.,
5 through a codec mode request (CMR) [9]) to adapt the FEC offset and the rate at which the partial redundancy is transmitted to improve the end user experience.

ACELP and TCX Partial Frame Decoding

The “bit-stream parser” module 98 in Figure 9 extracts the RF frame type information and
10 passes the partial copy information to the “partial frame decoding” module 97. Depending on the RF frame type, if the current frame corresponds to an ACELP partial copy, then the RF parameters (e.g., LSPs, ACB and/or FCB gains, and upper band gain) are decoded for ACELP synthesis. ACELP partial copy synthesis follows similar steps to that of the primary frame decoding 96 except that the missing parameters (e.g., certain gains and pitch lags are only
15 transmitted in alternate subframes) are extrapolated.

Furthermore, if the previous frame used a partial copy for synthesis, then a post-processing is performed in the current frame for a smoother evolution of LSPs and temporal gains. The post-processing is controlled based on the frame type (e.g., VOICED or UNVOICED) and spectral tilt estimated in the previous frame. If the current frame corresponds to a TCX partial copy,
20 then the RF parameters are used to perform a highly-guided concealment.

Subjective quality tests

Extensive testing of the EVS channel aware mode has been conducted via subjective ITU-T P.800 Mean Opinion Score (MOS) tests conducted at an independent test laboratory with 32
25 naïve listeners. The tests were conducted for both WB and SWB, using absolute category rating (ACR) and degradation category rating (DCR) test methodologies [24], respectively. Since the channel aware mode is specifically designed to improve performance for VoLTE networks, evaluating the performance in such networks is critical for establishing the potential benefits. Therefore, testing was conducted using codec outputs from simulations in which
30 VoLTE-like patterns of packet delays and losses were applied to received RTP packets before insertion into the de-jitter buffer. Four of these patterns – or, delay-loss profiles – were derived from real-world call logs of RTP packet arrival times collected in VoLTE networks in South Korea and the United States.

The resulting profiles mimic closely VoLTE network characteristics under different channel error conditions. In deriving the profiles, characteristics such as jitter, temporal evolution of jitter, and burstiness of errors were considered. These four profiles are identified in Figure 10 as profiles 7, 8, 9 and 10, and correspond to frame erasure rates (FER) at the decoder of approximately 3%, 6%, 8%, and 10%, respectively. These same four profiles have also been selected by 3GPP for use by that body for its own characterization testing of the EVS channel aware mode under channel impairments.

In addition to the VoLTE profiles, all codecs considered here were tested under error-free conditions and also for an HSPA profile included in the 3GPP MTSI specification [10] that yields about 6% frame erasure rate at the decoder. In all of the experiments, the EVS conditions used the reference EVS de-jitter buffer [13]. The AMR-WB conditions used a fixed delay buffer to convert delay-loss profiles to packet-loss profiles, such that packets experiencing a delay greater than a fixed threshold are discarded as described in the EVS performance requirements specification [14].

The ACR scores for the WB case are shown in Figure 10. For each profile, starting with the error-free ("Clean") profile, the chart compares (from left to right) AMR-WB, EVS AMR-WB IO mode, EVS baseline WB, and EVS WB channel aware ("RF"). The AMR-WB and EVS AMR-WB IO conditions used a higher bit rate of 15.85 kbps, whereas both EVS conditions used the same 13.2 kbps rate. These results show that the channel aware mode provides a statistically significant improvement compared to the non-channel-aware mode under all frame erasure conditions, even while maintaining equivalent quality under error-free conditions. Notably, the channel aware mode quality degrades much more gracefully even out to the 10% FER of profile 10. Compared to the AMR-WB and AMR-WB-IO conditions, the quality benefit is even more dramatic at these FER rates and has the potential to restore intelligibility under periods of high loss as might be encountered during a handoff, poor radio conditions, edge of the cell scenarios, or even on best-effort networks [8].

The performance advantage of the channel aware mode is similarly compelling in the super-wideband mode, the results for which are shown in Figure 11. As with WB, the channel aware mode does not degrade performance under error-free conditions, but has a statistically significant performance benefit under each of the lossy profiles, with the degree of improvement increasing as error rate increases. Figure 11 also shows the substantial improvement of EVS SWB Channel Aware mode at 13.2 kb/s compared to AMR-WB-IO at its maximum rate of 23.85 kb/s.

Conclusions

The Channel Aware coding mode of the new 3GPP EVS codec offers users and network operators a highly error resilient coding mode for VoLTE at a capacity operating point similar to the most widely used bit rates of existing deployed services based on AMR and AMR-WB.

5 The mode gives the codec the ability to maintain high quality WB and SWB conversational voice service even in the presence of high FER that may occur during network congestion, poor radio frequency coverage, handoffs, or in best-effort channels. Even with its graceful quality degradation under high loss, the impact to quality is negligible under low loss or even no-loss conditions. This error robustness offered by the Channel Aware mode further allows
10 for relaxing certain system level aspects such as frequency of re-transmissions and reducing scheduler delays. This in turn has potential benefits such as increased network capacity, reduced signaling overhead and power savings in mobile handsets. Use of the Channel Aware mode, therefore, can be beneficial in most networks without capacity impact to insure high quality communications.

15 Summarizing, the present invention utilizes the fact that the coder knows about the channel quality, for improving the speech/audio quality under erroneous conditions. In contrast to state of the art channel aware coding, the idea is to not have a partial copy that is just a low bitrate version of the primary encoded frame, but the partial copy consist of multiple key parameters that will enhance drastically the concealment. Therefore the decoder needs to distinguish
20 between regular concealment mode where all parameters are concealed and frameless mode where the partial copy parameters are available. Special care need to be taken for burst frameless for cases where the concealment needs to switch between partial and full concealment.

While this invention has been described in terms of several embodiments, there are alterations, permutations, and equivalents which fall within the scope of this invention. It should also be
25 noted that there are many alternative ways of implementing the methods and compositions of the present invention. It is therefore intended that the following appended claims be interpreted as including all such alterations, permutations and equivalents as fall within the true spirit and scope of the present invention.

30 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. Some or all of the method steps may be
35 executed by (or using) a hardware apparatus, like for example, a microprocessor, a

programmable computer or an electronic circuit. In some embodiments, some one or more of the most important method steps may be executed by such an apparatus.

The inventive encoded audio signal can be stored on a digital storage medium or can be transmitted on a transmission medium such as a wireless transmission medium or a wired transmission medium such as the Internet.

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 Blu-Ray, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed. Therefore, the digital storage medium may be computer readable.

Some embodiments according to the invention comprise a data carrier having electronically readable control signals, which are capable of cooperating with a 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 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.

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.

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. The data carrier, the digital storage medium or the recorded medium are typically tangible and/or non-transitionary.

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.

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
5 for performing one of the methods described herein.

A further embodiment according to the invention comprises an apparatus or a system configured to transfer (for example, electronically or optically) a computer program for performing one of the methods described herein to a receiver. The receiver may, for example, be a computer, a mobile device, a memory device or the like. The apparatus or system may,
10 for example, comprise a file server for transferring the computer program to the receiver .

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
15 methods are preferably performed by any hardware apparatus.

The apparatus described herein may be implemented using a hardware apparatus, or using a computer, or using a combination of a hardware apparatus and a computer.

The methods described herein may be performed using a hardware apparatus, or using a computer, or using a combination of a hardware apparatus and a computer.

20 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 herein.

References

- [1] RTP Payload for Redundant Audio Data", Internet Engineering Task Force, RFC 2198, September 1997
- [2] US 6,757,654 - "Forward error correction in speech coding", Westerlund, M. and al., 29 June 2004
- [3] "Adaptive joint playout buffer and FEC adjustment for Internet telephony" C. Boutremans, J.-Y. Le Boudec, INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies; 04/2003
- [4] Patent application: AUDIO DECODER AND METHOD FOR PROVIDING A DECODED AUDIO INFORMATION USING AN ERROR CONCEALMENT BASED ON A TIME DOMAIN EXCITATION SIGNAL
- [5] Patent application: AUDIO DECODER AND METHOD FOR PROVIDING A DECODED AUDIO INFORMATION USING AN ERROR CONCEALMENT MODIFYING A TIME DOMAIN EXCITATION SIGNAL
- [6] 3GPP TS 26.448: "Codec for Enhanced Voice Services (EVS); Jitter Buffer Management".
- [7] 3GPP TS 26.442: "Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point)".
- [8] D. J. Sinder, I. Varga, V. Krishnan, V. Rajendran and S. Villette, "Recent Speech Coding Technologies and Standards," in Speech and Audio Processing for Coding, Enhancement and Recognition, T. Ogunfunmi, R. Togneri, M. Narasimha, Eds., Springer, 2014.
- [9] J. Sjöberg, M. Westerlund, A. Lakaniemi and Q. Xie, "RTP Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs," April 2007. [Online]. Available: <http://tools.ietf.org/html/rfc4867>.
- [10] 3GPP TS 26.114, "Multimedia Telephony Service for IMS," V12.7.0, September 2014.
- [11] 3GPP TS 26.445: "EVS Codec Detailed Algorithmic Description; 3GPP Technical Specification (Release 12)," 2014.

- [12] 3GPP, TS 26.447, "Codec for Enhanced Voice Services (EVS); Error Concealment of Lost Packets (Release 12)," 2014.
- [13] 3GPP TS 26.448: "EVS Codec Jitter Buffer Management (Release 12)," 2014.
- [14] 3GPP Tdoc S4-130522, "EVS Permanent Document (EVS-3): EVS performance requirements," Version 1.4.
- [15] S. Bruhn, et al., "Standardization of the new EVS Codec," submitted to IEEE ICASSP, Brisbane, Australia, April, 2015.
- [16] M. Dietz, et al., "Overview of the EVS codec architecture," submitted to IEEE ICASSP, Brisbane, Australia, April, 2015.
- [17] V. Atti, et al., "Super-wideband bandwidth extension for speech in the 3GPP EVS codec," submitted to IEEE ICASSP, Brisbane, Australia, April, 2015.
- [18] G. Fuchs, et al., "Low delay LPC and MDCT-based Audio Coding in EVS," submitted to IEEE ICASSP, Brisbane, Australia, April, 2015.
- [19] S. Disch et al., "Temporal tile shaping for spectral gap filling within TCX in EVS Codec," submitted to IEEE ICASSP, Brisbane, Australia, April, 2015.
- [20] J. Lecomte et al., "Packet Loss Concealment Technology Advances in EVS," submitted to IEEE ICASSP, Brisbane, Australia, April, 2015.
- [21] B. Bessette, et al, "The adaptive multi-rate wideband speech codec (AMR-WB)," IEEE Trans. on Speech and Audio Processing, vol. 10, no. 8, pp. 620-636, November 2002.
- [22] E. Ravelli, et al., "Open loop switching decision based on evaluation of coding distortions for audio codecs," submitted to IEEE ICASSP, Brisbane, Australia, April, 2015.
- [23] M. Jelínek, T. Vaillancourt, and Jon Gibbs, "G.718: A New Embedded Speech and Audio Coding Standard with High Resilience to Error-Prone Transmission Channels," IEEE Communications Magazine, vol. 47, no. 10, pp. 117-123, October 2009.
- [24] ITU-T P.800, "Methods for Subjective Determination of Transmission Quality," International Telecommunication Union (ITU), Series P., August 1996.

CLAIMS

1. An encoder (1) for coding speech-like content and/or general audio content (2),
wherein the encoder (1) is configured to embed, at least in some frames (8), parameters
5 (6) in a bitstream (5), which parameters (6) enhance a concealment in case an original
frame (4) is lost, corrupted or delayed.
2. The encoder according to claim 1, wherein the encoder (1) is configured to create a
primary frame (4) and a partial copy (8), wherein the partial copy (8) is not a low bitrate
10 version of the primary frame (4) but wherein the partial copy (8) contains the parameters
(6).
3. The encoder according to claim 1 or 2, wherein the encoder (1) is configured to delay
the parameters (6) by some time and to embed the parameters (6) in a packet (9) which
is encoded and sent later in time.
4. The encoder according to one of claims 1 to 3, wherein the encoder (1) is configured
15 to reduce a primary frame bitrate, wherein the primary frame bitrate reduction and a
partial copy frame coding mechanism together determine a bitrate allocation between
the primary frames (4) and the partial copy frames (8) to be included within a constant
total bitrate.
5. The encoder according to one of claims 2 to 4, wherein the encoder (1) is configured
20 to create a primary frame (4) of one of the speech-like content type and the general
audio content type in combination with a partial copy (8) of the other one of the speech-
like content type and the general audio content type.
6. The encoder according to one of the preceding claims, wherein the encoder (1) is part
of a codec using a TCX coding scheme.
- 25 7. The encoder according to claim 6, wherein the encoder (1) is configured to detect
whether the frame contains a noisy audio signal or whether the frame contains a noise
floor with sharp spectral lines that are stationary over a period of time, and to embed,
based on the detection, the parameters (6) into a TCX frame.
8. The encoder according to one of claims 6 or 7, wherein the parameters (6) comprise
30 ISF or LSF parameters, in particular predictively coded ISF or LSF parameters.

9. The encoder according to one of claims 6 to 8, wherein the parameters (6) comprise signal classification parameters.
10. The encoder according to one of claims 6 to 9, wherein the parameters (6) comprise a TCX global gain or a TCX global level.
- 5 11. The encoder according to one of claims 6 to 10, wherein the parameters (6) comprise at least one of a window information and a spectral peak position.
12. The encoder according to claims 1 to 5, wherein the encoder (1) is part of a switched codec, wherein the switched codec consists of at least two core coding schemes, wherein a first core coding scheme uses ACELP and a second core coding scheme
10 uses TCX.
13. The encoder according to claim 12, wherein the encoder (1) is configured to not put a partial copy (8) on top of a TCX frame (4) after a switch when there is a first TCX frame after an ACELP frame.
14. The encoder according to one of claims 6 to 13, wherein the encoder (1) is configured
15 to analyze the signal before encoding and to turn off the partial copy usage or to provide a reduced partial copy (8) based on the analyzed signal.
15. The encoder according to claims 1 to 14, wherein the encoder (1) is configured to choose between multiple partial copy modes which use different amounts of information and/or different parameter sets (6), wherein the selection of the partial copy mode is
20 based on parameters.
16. The encoder according to claim 15, wherein at least one of the multiple partial copy modes is a frequency domain concealment mode.
17. The encoder according to claim 15, wherein at least two of the multiple partial copy modes are different time domain concealment modes.
- 25 18. The encoder according to claim 17, wherein one of the at least two time domain concealment modes is selected if a frame contains a transient or if a global gain of the frame is lower than a global gain of a previous frame.
19. The encoder according to one of claims 1 to 18, wherein the encoder (1) is configured to send a LTP lag if LTP data is present.
- 30 20. The encoder according to one of claims 1 to 19, wherein the encoder (1) is configured to send a classifier information.

21. The encoder according to one of claims 1 to 20, wherein the encoder (1) is configured to send at least one of LPC parameters, LTP Gain, Noise Level and Pulse Position.
22. A decoder (31) for decoding speech-like content and/or general audio content (2),
wherein the decoder (31) is configured to use parameters (6) which are sent
5 later in time to enhance a concealment in case an original frame (4) is lost, corrupted or delayed.
23. The decoder according to claim 22, wherein the decoder (31) is configured to receive a primary frame (4) and a partial copy (8), wherein the partial copy (8) is not a low
10 bitrate version of the primary frame (4) but wherein the partial copy (8) contains the parameters (6).
24. The decoder (31) according to one of claims 22 or 23, wherein the parameters (6) are contained in a partial copy (8), and wherein the decoder (31) is configured to receive from a de-jitter buffer the partial copy (8b) of a currently lost frame (4b) if it is available.
25. The decoder according to one of claims 22 to 24, wherein the decoder (31) is configured
15 to receive a primary frame (4) of one of the speech-like content type and the general audio content type in combination with a partial copy (8) of the other one of the speech-like content type and the general audio content type
26. The decoder according to one of claims 22 to 25, wherein the decoder (31) is part of a codec using a TCX codec scheme.
- 20 27. The decoder according to claim 26, wherein the parameters (6) comprise ISF or LSF parameters, in particular predictively coded ISF or LSF parameters.
28. The decoder according to one of claims 26 or 27, wherein the parameters (6) comprise signal classification parameters.
29. The decoder according to one of claims 26 to 28, wherein the parameters (6) comprise
25 a TCX global gain or a TCX global level.
30. The decoder according to one of claims 26 to 29, wherein the parameters (6) comprise at least one of a window information and a spectral peak position.
31. The decoder according to claims 22 to 25, wherein the decoder (31) is part of a switched
30 codec, wherein the switched codec consists of at least two core coding schemes, wherein a first core coding scheme uses ACELP and a second core coding scheme uses TCX.

32. The decoder according to claim 31, wherein the decoder (31) is configured to use, after a switch, ACELP concealment in the case that a first TCX frame after an ACELP frame is not available to the decoder (31).
- 5 33. The decoder according to claims 22 to 32, wherein the decoder (31) is configured to choose between multiple partial copy modes which use different amounts of information and/or different parameter sets (6).
34. The decoder according to claim 33, wherein at least one of the multiple partial copy modes is a frequency domain concealment mode.
- 10 35. The decoder according to claim 33, wherein at least two of the multiple partial copy modes are different time domain concealment modes.
36. The decoder according to one of claims 22 to 35, wherein the decoder (31) is configured to receive a LTP lag if LTP data is present.
37. The decoder according to one of claims 22 to 36, wherein the decoder (31) is configured to receive a classifier information.
- 15 38. The decoder according to one of claims 22 to 37, wherein the decoder (31) is configured to receive at least one of LPC parameters, LTP Gain, Noise Level and Pulse Position.
39. The decoder according to one of claims 22 to 38, wherein the decoder (31) is configured to decrease a pitch gain and a code gain with two different factors in dependence on a concealment mode.
- 20 40. The decoder according to claim 39, wherein a first factor is 0.4 and a second factor is 0.7.
41. The decoder according to one of claims 23 to 40, wherein the decoder (31) is configured to not take into account a pitch decoded from the partial copy (8b) if the previous primary frame (4a) is lost, and wherein the decoder (31) is configured to fix the pitch to a predicted pitch for the following lost primary frame instead of using the pitch transmitted.
- 25 42. A system comprising an encoder (1) according to one of claims 1 to 21 and a decoder (31) according to one of claims 22 to 41.
43. A method for encoding speech-like content and/or general audio content (2), the method comprising the step of:
- 30

embedding, at least in some frames (8), parameters (6) in a bitstream (5), which parameters (6) enhance a concealment in case an original frame (4) is lost, corrupted or delayed.

44. A method for decoding speech-like content and/or general audio content, the method comprising the step of:

using parameters (6) which are sent later in time to enhance a concealment in case an original frame (4) is lost, corrupted or delayed

45. A computer program for implementing the method of one of claims 43 or 44 when being executed on a computer or signal processor.

46. An encoder (1) for coding audio content (2), wherein the encoder (1) is configured to

provide a primary encoded representation (4) of a current frame and an encoded representation (8) of at least one error concealment parameter (6) for enhancing a decoder-sided error concealment of the current frame, wherein

the encoder (1) is configured to select the at least one concealment parameter (6) based on one or more parameters representing a signal characteristic of the audio content (2) contained in the current frame.

47. The encoder (1) according to claim 46, wherein the decoder-sided error concealment is an extrapolation-based error concealment.

48. The encoder according to one of claims 46 or 47, wherein the encoder (1) is configured to combine the encoded representation (8b) of the at least one error concealment parameter (6) of the current frame with a primary encoded representation (4c) of a future frame into a transport packet (9) such that the encoded representation (8b) of the at least one error concealment parameter (6) of the current frame is sent with a time delay relative to the primary encoded representation (4b) of the current frame.

49. The encoder according to one of claims 46 to 48, wherein the encoder (1) is configured to selectively choose between at least two modes for providing an encoded representation (8) of the at least one error concealment parameter (6).

50. The encoder according to claim 49, wherein the selection of a mode for providing an encoded representation (8) of the at least one error concealment parameter (6) is based on parameters which comprise at least one of a frame class, a LTP pitch, a LTP gain

and a mode for providing an encoded representation (8) of the at least one error concealment parameter (6) of one or more preceding frames.

51. The encoder according to one of claims 49 or 50, wherein at least one of the modes for providing an encoded representation (8) of the at least one error concealment parameter (6) is a time domain concealment mode such that the encoded representation (8) of the at least one error concealment parameter (6) comprises one or more of a TCX LTP lag and a classifier information.

52. The encoder according to one of claims 49 to 51, wherein at least one of the modes for providing an encoded representation (8) of the at least one error concealment parameter (6) is a time domain concealment mode that is selected if the audio content (2) contained in the current frame contains a transient or if the global gain of the audio content (2) contained in the current frame is lower than the global gain of the preceding frame.

53. The encoder according to one of claims 49 or 50, wherein at least one of the modes for providing an encoded representation (8) of the at least one error concealment parameter (6) is a frequency domain concealment mode such that the encoded representation (8) of the at least one error concealment parameter (6) comprises one or more of an LSF parameter, a TCX global gain and a classifier information.

54. The encoder according to one of claims 46 to 53, wherein the encoder (1) uses at least a TCX coding scheme.

55. A decoder (31) for decoding audio content (2), wherein the decoder (31) is configured to

receive a primary encoded representation (4) of a current frame and/or an encoded representation (8) of at least one error concealment parameter (6) for enhancing a decoder-sided error concealment of the current frame, wherein

the decoder (31) is configured to use the error concealment for at least partly reconstructing the audio content (2) of the current frame by using the at least one error concealment parameter (6) in the case that the primary encoded representation (4) of the current frame is lost, corrupted or delayed.

56. The decoder (31) according to claim 55, wherein the decoder-sided error concealment is an extrapolation-based error concealment.

57. The decoder according to claim 55 or 56, wherein the decoder (31) is configured to extract the error concealment parameter (6) of a current frame from a packet (9b) that is separated from a packet (9a) in which the primary encoded representation (4) of the current frame is contained
- 5 58. The decoder according to one of claims 55 to 57, wherein the decoder (31) is configured to selectively choose between at least two error concealment modes which use different encoded representations (8) of one or more error concealment parameters (6) for at least partially reconstructing the audio content (2) using the error concealment.
- 10 59. The decoder according to claim 58, wherein at least one of the at least two error concealment modes which uses different encoded representations (8) of one or more error concealment parameters (6) is a time domain concealment mode wherein the encoded representation (8) of the at least one error concealment parameter (6) comprises at least one of a TCX LTP lag and a classifier information.
- 15 60. The decoder according to one of claims 58 or 59, wherein at least one of the at least two error concealment modes which uses different encoded representations (8) of one or more error concealment parameters (6) is a frequency domain concealment mode wherein the encoded representation (8) of the at least one error concealment parameter (6) comprises one or more of an LSF parameter, a TCX global gain and a classifier information.
- 20 61. The decoder according to one of claims 55 to 60, wherein the decoder (31) uses at least a TCX coding scheme.
62. Apparatus for error concealment, the apparatus being configured for performing a standard concealment mechanism for a lost frame (4) and to use transmittable parameters (6) to enhance the concealment.
- 25 63. Apparatus for error concealment, the apparatus being configured for not having a partial copy that is just a low bitrate version of the primary (4), but for having a partial copy (8) consisting of multiple key parameters (6) for enhancing the concealment.
64. Apparatus for error concealment having a receiver (90) comprising a de-jitter buffer (95) for providing a partial redundant copy (99) of a current lost frame (92) if it is available
30 in any of the future frames (94), wherein the apparatus is configured for reading a partial redundant information bitstream and for updating corresponding parameters.

65. Switched Coder (1) or decoder (31), in which there are two or more core coding schemes, whereas for example one uses ACELP for coding speech-like content and the second use TCX for coding general audio content wherein ACELP frames are processed using a partial redundant copy coding and TCX frames are processed using a different approach, wherein in frames that are close to a core coder switch, two special cases can occur, namely:

ACELP primary frame with partial copy generated from future TCX frame on top,
TCX primary frame with partial copy generated from future ACELP frame on top

wherein, for these cases, both core coders are configurable to create primary frames in combination with partial copies from the other coder type, without infringing the required total size of a frame, to assure a constant bitrate, or wherein:

a first TCX frame after an ACELP frame, where, if this frame gets lost and thus is not available to the decoder (31), the proposed technique will TCX conceal the frame using partial copy information (8) that has been transported in top of another frame (4), wherein concealment needs a preceding frame for extrapolating the signal content, ACELP concealment is used (as the previous frame was ACELP) and wherein it is decided already in the encoder, to not put a partial copy (8) on top of a TCX frame after a switch, or

where there is a signal-adaptive partial copy selection, where a signal is analyzed before encoding to determine if the usage of partial copy is favorable, wherein if the signal could be concealed satisfyingly well without the help of additional partial copy info within the decoder (31), but the clean channel performance suffers because of reduced primary frame (4), a partial copy usage is turned off or a specifically reduced partial copy (8) is used within the encoder (1).

66. Transform Domain Coder (1) or decoder (31), wherein an en-/decoding scheme is used, where at least in some frames (8) redundant coding parameters (6) are embedded in the bitstream (5) and transmitted to the decoder side or wherein a redundant info (6) is delayed by some time and embedded in a packet (9) which is encoded and sent later in time such that the info (6) can be used in the case of the decoder (31) already having the future frame available, and the original frame (4) is lost, corrupted or delayed even more.

67. Transform domain coder (1) or decoder (31) as before, wherein redundant information (6) comprises ISF/LSF parameters:

ISF/LSF parameter representation is used for quantization and coding of LPC parameters. In TCX the LPC is used to represent the masking threshold. This is an essential parameter and very helpful to have available correctly on decoder side in case of a frame loss. Especially if the ISF/LSFs are coded predictively the concealment quality will improve significantly by having this info available during concealment, because the predictor states on decoder side will stay correct (in sync to encoder) and this will lead to a very quick recovery after the loss;

Signal classification:

Signal classification is used for signaling the content types: UNVOICED, UNVOICED TRANSITION, VOICED TRANSITION, VOICED and ONSET. Typically this type of classification is used in speech coding and indicating if tonal/predictive components are present in the signal or if the tonal/predictive components are changing. Having this information available on the decoder side during concealment may help to determine the predictability of the signal and thus it can help adjusting the amplitude fade-out speed, the interpolation speed of the LPC parameters;

TCX global gain/level:

The global gain may be transmitted to easily set the energy of the concealed frame to the correct (encoder determined level) in case it is available;

Window information like overlap length; or

Spectral peak positions to help tonal concealment.

68. Method or computer program similar to an apparatus as claimed above.