

Audio Decoder Supporting a Set of Different Loss Concealment Tools

Description

- 5 The present application is concerned with an audio decoder supporting a set of different loss concealment tools.

Packet loss concealment (PLC) is used in audio codecs to conceal lost or corrupted packets during the transmission from the encoder to the decoder. PLC is performed at the
10 decoder side and works by extrapolating the decoded signal either in the transform-domain or in the time-domain. Ideally, the concealed signal should be artifact-free and should have the same spectral characteristics as the missing signal.

Error robust audio codecs, as described in [2] and [4], have generally multiple
15 concealment methods for the various signal types like speech as an example for a monophonic signal, music as an example for polyphonic signal or noise signal. The selection is based on a set of signal features, which are either transmitted and decoded from the bit stream or estimated in the decoder.

20 Pitch-based PLC techniques generally produce good results for speech and monophonic signals. These approaches assume that the signal is locally stationary and recover the lost signal by synthesizing a periodic signal using an extrapolated pitch period. These techniques are widely used in CELP-based speech coding such as in ITU-T G.718 [2]. They can also be used for PCM coding such as in ITU-T G.711 [3] and more recently they
25 were applied to DECT-based audio coding, the best example being TCX time domain concealment, TCX TD-PLC, in the 3GPP EVS standard [4].

The pitch-lag is the main parameter used in pitch-based PLC. This parameter can be estimated at the encoder-side and encoded into the bit stream. In this case, the pitch-lag
30 of the last good frame is used to conceal the current lost frame such as in [2] and [4]. If there is no pitch-lag in the bitstream, it can be estimated at the decoder-side by running a pitch detection algorithm on the decoded signal such as in [3].

For non-periodic, non-tonal, noise-like signals, a low complexity technique called frame
35 repetition with sign scrambling has been found to be effective. It is based on repeating the last frame and multiplying the spectral coefficients with a randomly generated sign to

conceal the lost frame. One example of MDCT frame repetition with sign scrambling can be found in the 3GPP EVS standard [4].

5 For tonal polyphonic signals or complex music signals a method is used which is based on predicting the phase of the spectral coefficients of any detected tonal component. This method shows a consistent improvement for stationary tonal signals. A tonal component consists of a peak that also existed in the previous received frame(s). The phase of the spectral coefficients belonging to the tonal components is determined from the power spectrum of the last received frame(s). One example of tonal MDCT concealment can be
10 found in the 3GPP EVS standard [4].

Summarizing the above, different PLC methods are known but they are specific for certain situations, i.e., for certain audio characteristics. That is, an audio coder supporting several of these PLC methods should have a mechanism to choose the most suitable PLC
15 method at the time of encountering frame or packet loss. The most suitable PLC method is the one leading to the least noticeable substitute for the lost signal.

The object of the present application is to provide an audio decoding concept which allows an audio decoder which comprises a set of different loss concealment tools to achieve an
20 improved loss concealment.

This object is achieved by the subject matter of the independent claims of the present application.

25 The idea of the present invention is based on the finding that an assignment of one of phase set of different loss concealment tools of an audio decoder to a portion of the audio signal to be decoded from a data stream, which portion is affected by loss, that is the selection out of the set of different loss concealment tools, may be made in a manner leading to a more pleasant loss concealment if the assignment/selection is done based on
30 two measures: A first measure which is determined measures a spectral position of a spectral centroid of a spectrum of the audio signal and a second measure which is determined measures a temporal predictability of the audio signal. The assigned or selected loss concealment tool may then be used to recover the portion of the audio signal.

35

For instance, based on the aforementioned first and second measures, one of first and second loss concealment tools may be assigned to the lost portion with a first being configured to recover the audio signal by audio signal synthesis using a periodic signal of a periodicity which depends on a pitch value derived from the data stream, and the
5 second loss concealment tool may be configured to recover the audio signal by detecting tonal spectral components of the audio signal, performing phase detection at the tonal spectral components and audio signal synthesis by combining the signals of periodicities which depend on the tonal spectral components at adjustment of a mutual phase shift between the signals depending on the phase detection. In other words, based on the first
10 and second measures, one of a tonal frequency domain PLC tool and a tonal time domain PLC tool may be assigned to the lost portion.

In accordance with an embodiment, the assignment/selection for a lost portion is performed in stages: A third measure measuring a tonality of the spectrum of the audio
15 signal is determined and one of a first and second subsets of one or more loss concealment tools out of the set of different loss concealment tools is assigned to the lost portion, and merely if the first subset of one or more loss concealment tools is assigned to the lost portion, the assignment of the one PLC tool for the lost portion is performed based on the first and second measures out of this first subset. Otherwise, the
20 assignment/selection is performed out of the second subset.

According to an aspect of the present application, there is provided an audio decoder for decoding an audio signal from a data stream, the audio decoder comprising a set of different loss concealment tools and configured to determine a first measure measuring a
25 spectral position of a spectral centroid of a spectrum of the audio signal, determine a second measure measuring a temporal predictability of the audio signal, assign one of the set of different loss concealment tools to a portion of the audio signal affected by loss based on the first and second measures, and recover the portion of the audio signal using the one loss concealment tool assigned to the portion, wherein the set of different loss
30 concealment tools comprises: a first loss concealment tool for audio signal recovery of monophonic portions, and a second loss concealment tool for audio signal recovery of polyphonic portions, and wherein the audio decoder is configured to, in assigning the one of the set of different loss concealment tools to the portion of the audio signal based on the first and second measures, assign the first loss concealment tool to the portion the
35 more likely the lower the spectral position of the spectral centroid is and the higher the temporal predictability is, and assign the second loss concealment tool to the portion the

more likely the higher the spectral position of the spectral centroid is and the lower the temporal predictability is.

According to another aspect of the present application, there is provided an audio decoder
5 for decoding an audio signal from a data stream, the audio decoder comprising a set of
different loss concealment tools and configured to: determine a first measure measuring a
spectral position of a spectral centroid of a spectrum of the audio signal, determine a
second measure measuring a temporal predictability of the audio signal, assign one of the
set of different loss concealment tools to a portion of the audio signal affected by loss
10 based on the first and second measures, and recover the portion of the audio signal using
the one loss concealment tool assigned to the portion, wherein the audio decoder is
configured to, in assigning one of the set of different loss concealment tools to a portion of
the audio signal affected by loss based on the first and second measures, perform a
summation over the first and second measures so as to obtain a scalar sum value and
15 subjecting the scalar sum value to thresholding.

According to another aspect of the present application, there is provided a method for
performing loss concealment in audio decoding an audio signal from a data stream, the
method comprising: determining a first measure measuring a spectral position of a
20 spectral centroid of a spectrum of the audio signal, determining a second measure
measuring a temporal predictability of the audio signal, assigning one of a set of different
loss concealment tools to a portion of the audio signal affected by loss based on the first
and second measures, and recovering the portion of the audio signal using the one loss
concealment tool assigned to the portion, wherein the set of different loss concealment
25 tools comprises: a first loss concealment tool for audio signal recovery of monophonic
portions, and a second loss concealment tool for audio signal recovery of polyphonic
portions, and wherein the assigning the one of the set of different loss concealment tools
to the portion of the audio signal based on the first and second measures comprises
assigning the first loss concealment tool to the portion the more likely the lower the
30 spectral position of the spectral centroid is and the higher the temporal predictability is,
and assigning the second loss concealment tool to the portion the more likely the higher
the spectral position of the spectral centroid is and the lower the temporal predictability is.

According to yet another aspect of the present application, there is provided a method for
35 performing loss concealment in audio decoding an audio signal from a data stream, the
method comprising: determining a first measure measuring a spectral position of a spectral

centroid of a spectrum of the audio signal, determining a second measure measuring a temporal predictability of the audio signal, assigning one of a set of different loss concealment tools to a portion of the audio signal affected by loss based on the first and second measures, and recovering the portion of the audio signal using the one loss concealment tool assigned to the portion, wherein the assigning one of the set of different loss concealment tools to a portion of the audio signal affected by loss based on the first and second measures comprises performing a summation over the first and second measures so as to obtain a scalar sum value and subjecting the scalar sum value to thresholding.

10

Advantageous aspects of the present application are the subject of dependent claims. Preferred embodiments of the present application are set out below with respect to the figures among which:

15 Fig. 1 shows in a schematic manner a block diagram of an audio decoder in accordance with an embodiment of the present application;

Fig. 2 shows a flow diagram of a mode of operation of the assigner 32 of the audio decoder of Fig. 1 in accordance with an embodiment;

20

Fig. 3 shows a schematic diagram illustrating a spectral centroid detection for determining the first measure for PLC assignment in Fig. 2 in accordance with an embodiment;

Fig. 4 shows a schematic diagram illustrating the temporal predictability detection for determining the second measure for PLC assignment of Fig. 2 in accordance with an embodiment;

5 Fig. 5 illustrates a schematic diagram for illustrating the mode of operation of PLC 1 of Fig. 2;

Fig. 6 shows a schematic diagram illustrating a mode of operation of PLC 2 in accordance with an embodiment;

10

Fig. 7 shows a flow diagram of a modified version of the assignment operation of Fig. 2, extended so as to perform an assignment out of a set of PLC tools for packet loss concealment PLC 1 to PLC 3 in accordance with an embodiment; and

15 Fig. 8 shows a decision tree for deciding among three PLC tools to which decision tree the embodiment of Fig. 7 corresponds.

Fig. 1 shows an audio decoder in accordance with an embodiment of the present application. As shown therein, the audio decoder 10 is configured to decode an audio
20 signal 12 from a data stream 14. The audio signal 12 may be encoded into data stream 14 in accordance with any suitable audio codec such as a time-domain based audio codec or a frequency-domain audio codec. The audio signal 12 may be coded into data stream 14 in units of temporal portions 16 of frames 18, respectively. To be more precise, the audio signal 12 may be temporally sub-divided into overlapping or non-overlapping temporal
25 portions or intervals 16, each of which corresponds to a certain one of frames 18 which data stream 14 is sub-divided into. Each frame 18 encodes a corresponding temporal portion 16. For instance, a portion 18 may contain information on linear prediction coefficients describing a spectral envelope of the audio signal within the corresponding portion 16. Additionally, frame 18 may have encoded thereinto spectral coefficients
30 describing a spectrum of the audio signal 12 within portion 16 which is to be shaped, for instance, by audio decoder 10 according to the linear prediction coefficients contained in that frame. An overlap add process might also be applied by the audio decoder 10 in reconstructing the audio signal 12 from the data stream 14. Naturally, the possibilities would also apply with the examples presented herein merely serving for ease of
35 understanding.

The data stream 14 may be received by the audio decoder 10 in a packetized form, i.e., in units of packets. The sub-division of data stream 14 into frame 18 itself represents a kind of packetization, i.e., the frames 18 represent packets. Additionally, data stream 14 may be packed into packets of a transport stream or media file format, but this circumstance is not inspected in further detail here. Rather, it should suffice to state that the reception of the data stream 14 by audio decoder 10 is liable to data or signal loss, called packet loss in the following. That is, some continuous portion 20 of the data stream 14 might have got lost during transmission, thus not received by audio decoder 10, so that the corresponding portion is missing and not available for the audio decoder 10. As a consequence, audio decoder 10 misses information in the data stream 14 so as to reconstruct a portion 22 corresponding to portion 20. In other words, the audio decoder 10 is not able to reconstruct portion 22 from data stream 14 in accordance with a normal audio decoding process implemented, for instance, in an audio decoding core 24 of the audio decoder, as portion 20 of data stream 14 is missing. Rather, in order to deal with such missing portions 20, audio decoder 10 comprises a set 26 of PLC tools 28 so as to recover or synthesize the audio signal 12 within portion 22 by a substitute signal 30. The PLC tools 28 comprised by set 26 differs in their suitability for different audio signal characteristics. That is, the degree of annoyance when using a certain PLC tool for the recovery of a signal substitute 30 within a certain portion 22 of the audio signal 12 depends on the audio signal characteristic at that portion 22 and PLC tools 28 within set 26 show mutually different degrees of annoyance for a certain set of audio signal characteristics. Accordingly, audio decoder 10 comprises an assigner 32 which assigns one of the set 26 of packet loss concealment tools 28 to portion 22 of the audio signal 12 which is affected by a packet loss such as the lost portion 22 of data stream 14. The assigner 32 tries to assign the best PLC tool 28 to portion 22, namely the one which leads to the lowest annoyance.

Once the assigner 32 has assigned a certain PLC tool 28 to a lost portion 22 of the audio signal 12, the audio decoder 10 recovers this portion 22 of the audio signal using the assigned PLC tool 28, thereby substituting the audio signal 12 within this portion 22, as it would have been reconstructed from the audio data stream 14 if the corresponding data stream portion 22 would not have got lost, by a substitute signal 30 obtained using the PLC tool 28 assigned for portion 22 by assigner 32.

As already indicated above, the assignment of a particular PLC tool 28 to a certain lost portion 22 should be made signal dependent in order to render the lost concealment as least annoying as possible. Signal dependency, however, is restricted to portions of data

stream 14 preceding the lost data stream portion 20 and, in accordance with the embodiment described herein, the assigner 32 acts as follows.

5 In order to explain this in more detail, reference is made to Fig. 2. Fig. 2 shows that the whole assignment process for a certain missing portion 22 might possibly be triggered by a loss detector 34 possibly present in audio decoder 10. In particular, if lost detection 36 performed by loss detector 34 reveals that some portion 20 of data stream 14 is missing or lost, as checked at 38, the following assignment process is triggered. A determination 40 is performed in order to determine a first measure 42 which measures a spectral position of a spectral centroid of a spectrum of the audio signal. That is, assigner 32 determines a spectral position of a center of mass of a spectrum of the audio signal, see Fig. 3. The audio decoder retrieves from a portion 44 preceding the lost portion 20 of data stream 14, preceding in data stream order, a spectrum 46 of the audio signal. As described above with respect to Fig. 1, it might be that data stream 14 has audio signal 12 encoded therein in spectral domain anyway so that no spectral decomposition is necessary for assigner 32 to obtain spectrum 46. For instance, spectral coefficients of a most recently received frame 18 or more than one most recently retrieved frame 18 of data stream 14, prior to the lost portion 20, is used to obtain spectrum 46. If more than one frame 18 is used, it could be that spectrum 46 used by assigner 32 is obtained by averaging. In order to perform the determination 40, the center of mass of this spectrum 46 is determined, i.e., a measure 42 measuring a spectral position 48 of spectrum 46. Later on, a specific example is presented.

Further, the assignment process triggered by loss detection comprises a determination 50 of a temporal predictability of the audio signal so as to obtain a measure 52 of this temporal predictability, see Fig. 4 for details. As shown therein, in accordance with an embodiment, the temporal predictability detection 50 may rely on the decoded signal or the audio signal 12 as derived from the data stream 14 up to the signal portion 22 which is missing owing to the loss of portion 20 of data stream 14. In other words, the temporal predictability detection 50 may be based on the audio signal 12 within portion 50 which immediately precedes lost portion 22, the loss of which is to be concealed, and which is decoded from portion 44 of data stream 14. The temporal predictability detection 50 may be done in a manner so that measure 52 is a measure for a self-similarity or autocorrelation of the signal portion 50 as illustrated at 54 in Fig. 4. The mutual shift s for which the self-similarity of signal 50 is measured by measure 52 may be determined by assigner 32 in different manners. For instance, assigner 32 may inspect a corresponding

pitch parameter conveyed in one or more of the frames 18 within portion 44 preceding lost portion 20 of data stream 14. That is, the mutual shift s at which the self-similarity may be measured may correspond to a pitch period with a pitch being determined based on a parameter in data stream 14, namely portion 44. The self-similarity or correlation 56 at
5 that pitch period shift may be used as the second measure 52.

It is obvious that the order of performing determination 40 and 50, respectively, may be switched or that both detections may be performed concurrently. Based on measures 42 and 52, an assignment 60 is performed. This assignment 60 selects one of two PLC tools
10 28 for concealment of loss of portion 22. This PLC tool, i.e., the assigned one 62, is then used for the concealment of the loss of portion 22.

As a brief note, it should be noted that the number of PLC tools 28, between which the selection by assignment 60 is performed, may be greater than two.

15 In accordance with an embodiment further outlined below, however, the PLC tool PLC 1 of Fig. 2 may be described as 1 using which the substitute signal 30, i.e., the audio signal estimate within portion 22, is obtained or recovered using tonal time domain packet loss concealment. In other words, PLC 1 may be a packet loss concealment dedicated for
20 audio signal recovery of monophonic portions. PLC 1 may recover an audio signal within a missing portion 22 of an audio signal 12 using a periodic signal of a periodicity which depends on a pitch parameter or pitch value derived from the data stream, namely from portion 44 of data stream 14, i.e., the portion 44 preceding the lost portion 20 of data stream 14.

25 The second PLC tool PLC 2 may be dedicated for the recovery of audio signals of polyphonic type. The concealment of this second PLC tool PLC 2 may be based on tonal frequency domain packet loss concealment.

30 With respect to Figs. 5 and 6, a possible implementation of PLC 1 and PLC 2 will be briefly explained.

Fig. 5 illustrates PLC 1. A pitch parameter or pitch value 66 conveyed in a frame 18 within portion 44 preceding lost portion 20 of data stream 14 is used to set a periodicity or period
35 length 68 of a periodic signal 70 which is then used to form a substitute or used to conceal the loss within portion 22 of audio signal 12. The pitch parameter or pitch value 66 may be

present in data stream 14 in order to be used by audio decoder 10 in case of normal audio decoding, i.e., no signal loss, for controlling, for instance, a harmonic filter tool or the like. That is, parameter 66 may be present in data stream 14 anyway. Otherwise, PLC tool 28 performing PLC 1 according to Fig. 5 could determine the pitch period 68 by analysis such as by analyzing the decoded signal 50 in front of the lost portion 22 or by analyzing the most recent accessible version of the spectrum such as spectrum 46 depicted in Fig. 3.

Fig. 6 illustrates PLC 2 in accordance with an embodiment. Here, the PLC tool 28, responsible for performing PLC 2, uses, for instance, one or two or more most recently obtained spectra as obtained from portion 44 of data stream 14 so as to detect or determine tonal spectral components therein, i.e., peaks 72 in the spectrum 46 or peaks 72 in spectrum 46 occurring at that position or a sufficiently similar position in the spectrum of a certain number of consecutive spectra or frames 18, respectively. Sufficiently similar positions may be ones the spectral distance of which is below a certain threshold. The spectral positions of the peaks 72 represent the tonal spectral components and here, at these spectral locations, the phase detection is performed by use of, for instance, or by evaluation of, for instance, a power spectrum of the audio signal. Then, within the temporal portion 22 within which the signal loss is to be concealed, a combination of signals, periodicities of which depend on the tonal spectral components, is formed so as to yield the supplement signal 30, wherein a mutual phase shift between the combined signals is adjusted depending on the phase detection. For instance, a phase is determined for each tonal component 72 or merely phase differences between these tonal components are determined, and a signal is formed as the substitute 30 within portion 22 which is synthesized by these tonal components 72 with obeying the phase differences or phases. The combination may be formed in spectral domain with deriving the substitute signal 30 by inverse transformation, or in time-domain directly by adding, for instance, appropriately mutually shifted signals, the mutual shift reflecting the mutual phase shift determined.

As described in more detail below, the assignment 60 may be done in a manner so that PLC 1 is chosen or assigned to portion 22 the more likely the lower the spectral position 48 is and the higher the temporal predictability is and, vice versa, PLC 2 is assigned or selected the more likely the higher the spectral position 48 is and the lower the temporal predictability is. A higher spectral position corresponds to a higher frequency and a lower spectral position to a lower frequency. By doing this in this manner, PLC 1 is more likely

chosen in case of portion 22 corresponding to lost speech and PLC 2 is more likely selected in case of portion 22 relating to polyphone signals or music.

For the sake of completeness, Fig. 7 shows the fact that the assignment process of Fig. 2 may be extended. In particular, as shown in Fig. 2, the assignment 60 has been done by restricting the assignment or selection of assignment 60 onto a subset 80 of PLC tools 28. The set 26 of PLC tools 28 may contain a further subset of one or more PLC tools 28, such as subset 82 and, when triggered by loss detection 36, a tonality detection 84 may be used by assigner 32 first in order to determine whether portion 22 relates to a tonal portion or not. The tonality determination 84 yields a tonality measure or indicator 86 and this tonality measure indicator 86 may be obtained in 84 by use of one or more parameters within portion 44 of data stream 14, for instance such as by inspecting whether or not a most recent frame 18 within portion 44 comprises a certain pitch parameter such as a pitch value 66 as described in Fig. 5. In case of absence, this may be interpreted as an indication that the audio signal is currently non-tonal and, in case of presence of the pitch parameter 66, this may be interpreted as indicating that the audio signal is currently tonal. This indication is then the tonality measure 86. Using measure 86 it is discriminated at 88 whether the audio signal is tonal or not and, if it is non-tonal, the PLC 62 assigned to portion 22 is assigned out of subset 82. Fig. 7 illustrates the case where one PLC tool 28 is comprised by subset 82 and this one is chosen. However, even here, a further selection out of subset 82 may follow. If tonality is confirmed at 88, the assignment 60 based on determinations 40 and 50 is performed with respect to subset 80 as described above with respect to Fig. 2.

PLC 3 may be a non-tonal PLC such as a PLC which recovers an audio signal for a portion 22 by use of frame repetition with or without replicate modification, when the replicate modification may, as indicated above, involve sign scrambling, i.e., a random sign flip of spectral coefficients of a most recently received spectrum such as spectrum 46 which is then inversely transformed and used to derive substitute signal 30.

Fig. 8 visualizes a decision tree for selecting among the PLC tools 28 in accordance with a manner which corresponds, for instance, to Fig. 7. Decisions A and B are made to decide which of three PLC methods PLC 1 to PLC 3, which are designed for different signal types, are to be used for a certain missing signal portion 22 in order to get the best concealment performance. The first decision, decision A, is based on tonality. If the signal turns out to be non-tonal, PLC 3 is used for concealment. If tonal, decision B is made.

Decision B checks the tonal signal characteristics based on measures 42 and 52 in the manner described above with respect to Figs. 2 and 7 and, depending on the characteristics, chooses one of PLC 1 or PLC 2. As explained above, PLC 1 may be a tonal time domain PLC for monophone and/or speech signals, while PLC 2 may be a tonal frequency domain PLC for a polyphone and/or music signals.

The decision tree of Fig. 8, thus, visualizes decisions, which may be taken between three PLC methods for different signal types to get the best concealment performance. The decision A, which may correspond to check 88 of Fig. 7, may be done by inspecting a signal type classification, i.e. by using it as an indication of, or by deriving therefrom a tonality indicator. The signal type classification is possibly present in each frame 18 and indicates a frame class of each frame. It could be calculated on encoder side and transmitted in the bitstream 14 to the audio decoder. Even alternatively, it could be calculated on decoder side. However, the calculation of the frame class is very complex and may require that all features are calculated in every frame due to a frame dependency of the frame class. Therefore, for low-complexity applications, it may be preferred if a simpler approach is used. As indicated above, the presence or absence of some pitch parameter 66 may be used as indicator 86.

The decision B which corresponds to assignment 60 based on determinations 40 and 52, yields a good choice between PLC#1 and PLC#2. In [6], such a choice has been done based on a stability measurement of the spectral envelope, which correlates to the short-term stationarity of the signal. However, the more stationary a signal is, the better the performance both tonal PLC methods PLC#1 and PLC#2 is. This means stationarity is, hence, not a suitable criterion to select the optimal tonal concealment method. The stationarity feature indicates tonality very well, however it cannot differentiate between speech/monophonic and polyphonic/music.

As discussed above, it is possible, to perform the decision tree of Fig. 8 using a PLC classifier represented by assigner 32, which may operate on a frame-by-frame basis without any inter-frame dependencies and thus necessitates merely a low complexity. It may calculate its classification features 42 and 52 only in case of a frame loss, as detected or checked at 38, and therefore does not add an immanent complexity offset in the error-free frames among frames 18.

The decision A may be done based on a tonality indicator 86, which can be the presence of a pitch value in the last good received audio frame. The decision B may be done by using the spectral centroid 48 and a long term prediction gain 56 calculated on the last good received audio frame.

5

The decision B may switch between a pitch-based time domain concealment method PLC#1, best suited for monophonically and speech-like signals, and frequency domain methods PLC#2, best suited for polyphone or complex music signals. An advantage of the classification of decision B results from the fact, that:

- 10 • the spectral centroid is spectrally located very low for speech signals and higher for music signals;
- the long term prediction gain is high for monophonic and stationary signals like speech signals and low for polyphonic or less stationary signals like tonal or complex music signals.

15

Therefore, a weighted combination of both features 42 and 52 may be used for decision B and assignment process 60 and results in a reliable discrimination of speech/monophonic and polyphonic/complex music signals. At the same time, the complexity may be kept low.

- 20 If the audio decoder receives a corrupted frame or if the frame is lost, i.e. encounters a lost portion 20, as detected at 38, the following may be done, wherein reference is also made to Fig. 2:

- 25 a. Decision A is made, at 88, whether to use a tonal concealment method like PLC#2 or PLC#1 for concealing the lost/corrupted frame representing portion 20, or concealing the loss of the corresponding portion 22. This decision A is based on a tonality indicator 68 for which the presence of a pitch value 66 in the last good frame in the bitstream can be used.
- 30 b. For a negative decision A, no tonal PLC is used. Instead, another PLC method is used in that case, i.e. PLC#3. Same may use frame repetition with sign scrambling.
- 35 c. For a positive decision A, one of two tonal PLC methods PLC#1 and PLC#2 is used to conceal the lost/corrupted frame. The two PLC methods can be either the time domain pitch-based PLC such as TCX TD-PLC of [4] or frequency domain

tonal concealment such as the tonal MDCT concealment in [4], wherein the corresponding description is incorporated herein by reference.

For a positive decision A, the features 42 and 52 may be calculated based on the last
5 good frame in the following manner:

- As feature or measure 52, a long term prediction gain $xcorr$:

$$xcorr = \frac{\sum_{k=0}^{N-1} x(k) \cdot x(k - T_c)}{\sqrt{(\sum_{k=0}^{N-1} x(k) \cdot x(k)) \cdot (\sum_{k=0}^{N-1} x(k - T_c) \cdot x(k - T_c))}}$$

may be computed in 50, where T_c is the pitch value of the last good frame and $x(k), k =$
10 $0 \dots N - 1$, are the last decoded time samples of the last good frame and

$$N = \begin{cases} T_c, & T_c < N_F \\ N_F, & else \end{cases}$$

where N_F can be a limited value like the maximum pitch value or a frame length (for example 10ms).

- 15 • As feature or measure 42, a spectral centroid sc :

$$sc = \frac{\sum_{k=0}^{N-1} |X_{s_lastGood}(k)| \cdot \frac{k+1}{N}}{\sum_{k=0}^{N-1} |X_{s_lastGood}(k)|}$$

may be computed in 40, where N is the length of the last received spectrum $X_{s_lastGood}(k)$
and $|X_{s_lastGood}(k)|$ means the magnitude spectrum.

- 20 The two calculated features are combined with the following formula:

$$class = w_1 \cdot xcorr + w_2 \cdot sc + \beta$$

where w_1 , w_2 and β are weights. In one embodiment, these are $w_1 = \frac{520}{1185}$, $w_2 = -1$ and
 $\beta = -\frac{287}{1185}$. Alternatives are setting w_1 , w_2 and β so that $\frac{1}{4} < w_1 < \frac{3}{4}$, $-2 < w_2 < -\frac{1}{2}$, and
 $-\frac{1}{2} < \beta < -\frac{1}{16}$. The weights may be normalized here to be in the range [-1:1]

25

Then, the PLC#1, e.g. time domain pitch-based PLC method, may be chosen if $class > 0$
in 60 and PLC#2, such as a frequency domain tonal concealment, otherwise.

Some notes shall be made with respect to the above description. For instance, the
30 spectrum, the spectral centroid of which is measured to obtain the first measure 42, might
be a so called weighted version such as a pre-emphasized version. Such weighting is

used, for instance, to adapt the quantization noise to the psychoacoustic masking threshold. In other words, it might be that the first measure 42 measuring a spectral position 48 of a spectral centroid of a psychoacoustic scaled spectrum of the audio signal. This might be especially advantageous in cases where the normal audio decoding coded

5 underlying audio decoding core 24 involves that data stream 14 has audio signal 12 encoded therein in spectral domain anyway, namely in the weighted domain. Additionally or alternatively, the spectrum, the spectral centroid of which is measured to obtain the first measure 42, is not necessarily one represented at a spectral resolution as high as the spectral resolution used in the audio decoding core 24 to transition to time

10 domain. Rather, it may be higher or lower. Even additionally or alternatively, it should be noted that the audio signal's spectrum also manifests itself in scale factors. Such scale factors might be transmitted in the data stream 14 along with spectral coefficients in order to, together, form a coded representation of the audio signal's spectrum. For a certain portion 22, the spectral coefficients are scaled according to the scale factors. There are

15 more spectral coefficients than scaler factors. Each scale factor, for instance, is assigned to one of several spectral bands, so called scale factor bands, into which the audio signal's bandwidth is partitioned. The scale factors, thus, define the spectrum of the audio signal for a certain portion in terms of envelope at some spectral resolution reduced compared to the one at which the quantized spectral coefficients are coded in the data

20 stream 14. It could even be that the spectral resolution at which the scale factors are coded in the data stream 14 is even lower than a spectral resolution at which the decoding core 24 performs the dequantization of the spectral coefficients. For instance, the decoding core 24 might subject the scale factors coded into the data stream 14 to spectral interpolation to obtain interpolated scale factors of higher spectral resolution as the ones

25 coded into the data stream, and use the interpolated scale factors for dequantization. Either one of the scale factors coded into the data stream and the interpolated scale factors might be used as the spectrum of the audio signal the spectral centroid of which is measured by the first measure 42. This means that centroid measurement becomes quite computational efficient to be determined as the number of computational operations to be

30 performed to determine the first measure is low compared to performing the centroid measurement at any higher resolution such as at the one at which the spectral coefficient are coded or some other resolution in case of obtaining the spectrum for the centroid measurement by subjecting the decoded audio signal to an extra spectral decomposition which would even further increase the efforts. Thus, as a concrete example, first and

35 second measures could be computed as follows based on coded down-sampled scale factors SNS (spectral noise shaping):

Firstly, a pitch value T_c might be computed as a basis:

$$T_c = \begin{cases} 0, & \text{pitch_present} = 0 \\ \text{pitch_int}, & \text{pitch_present} = 1 \end{cases} \quad (1)$$

where pitch_present and pitch_int are bitstream parameters derived by the decoder from
5 the last good frame. pitch_present can be interpreted as a tonality indicator.

As the second measure, a long term prediction gain $xcorr$ might be computed according to:

$$xcorr = \frac{\sum_{k=0}^{N-1} x(k) \cdot x(k - T_c)}{\sqrt{(\sum_{k=0}^{N-1} x(k) \cdot x(k)) \cdot (\sum_{k=0}^{N-1} x(k - T_c) \cdot x(k - T_c))}} \quad (2)$$

10 where $x(k), k = 0 \dots N - 1$ are the last decoded time samples and N can be can be a predetermined length value such as limited value like the maximum pitch value or a frame length N_F (for example 10ms), for example

$$N = \begin{cases} 2 \cdot \text{pitmin}, & T_c < 2 \cdot \text{pitmin} \\ T_c, & 2 \cdot \text{pitmin} \leq T_c \leq N_F \\ N_F, & T_c > N_F \end{cases} \quad (3)$$

where pitmin is the minimal pitch value. Thus, the second measure would be computed as
15 the self-similarity of the decoded audio time signal at the most recently received portion with itself, mutually shifted at the pitch.

As the second measure, a spectral centroid sc could be computed as:

$$sc = \frac{f_s}{48000} \cdot \frac{\sum_{k=0}^{15} g_d(k) \cdot \frac{\text{bands}(k)}{N_F}}{\sum_{k=0}^{15} g_d(k) \cdot (I_{f_s}(4k + 4) - I_{f_s}(4k))} \quad (4)$$

where f_s is the sampling rate and

$$20 \quad \text{bands}(k) = \sum_{b=I_{f_s}(4k)+1}^{I_{f_s}(4k+4)} b \quad \text{for } k = 0..15 \quad (5)$$

and I_{f_s} are non-uniform band indices, i.e. band indices defining for each band the lower and upper frequency border in a manner so that the band widths defined by the difference between the associated lower and upper border differ from each other such as increase with increasing frequency although the difference is optional. The band indices might be
25 defined in dependency of the sampling rate/frequency of the audio signal. Further,

$$g_d(k) = \frac{2^{scfQ_{-1}(k)}}{10^{k \cdot \frac{g_{tilt}}{150}}} \quad \text{for } k = 0..15 \quad (6)$$

where $scfQ_{-1}(k)$ is the scale factor vector stored in the bitstream of the last good frame and g_{tilt} is a predetermined tilt factor which might be set by default and, possibly,

depending on the sample frequency of the audio signal. The term $2^{scfQ-1(k)}$ is applied to get the scalefactors coded in the logarithmic domain back in the linear domain. The term $\frac{1}{10^{k \frac{g_{tilt}}{150}}}$ is applied to inverse the encoder side pre-emphasis filter, which is called de-emphasis filter.

- 5 The scale factor vector is calculated at encoder side and transmitted in the bitstream. It is determined on the energies per band of the MDCT coefficients, where the bands are non-uniform and follow the perceptually-relevant bark scale (smaller in low-frequencies, larger in high-frequencies). After smoothing, pre-emphasizing and transforming the energies to logarithmic domain, they are, at the encoder side, downsampled from 64 parameters to 16
- 10 parameters to form the scale factor vector, which afterwards is coded and transmitted in the bitstream. Thus, sc is a measure for a spectral position 48 of a spectral centroid of a spectrum 46 of the audio signal, here determined based on the spectrally coarse sampled version thereof, namely the SNS parameters.

- 15 The decision or selection among the various PLC methods may then be done with the criteria $xcorr$ and sc . Frame repetition with sign scrambling might be selected if $T_c = 0$ (which means that the tonality indicator $pitch_present = 0$). Otherwise, the value $class$ is calculated as follows:

$$class = \frac{7640}{32768} xcorr - sc - \frac{5112}{32768} \quad (7)$$

- 20 time domain pitch-based PLC method might be chosen if $class > 0$; frequency domain tonal concealment otherwise.

- Thus, an audio decoder for decoding an audio signal 12 from a data stream 14, which comprises a set 26 of different loss concealment tools 28 might be configured to
- 25 determine 40 a first measure 42 measuring a spectral position 48 of a spectral centroid of a spectrum 46 of the audio signal by deriving the spectrum from scale factors in a most recent non-lost portion of the data stream, determine 50 a second measure 52 measuring a temporal predictability of the audio signal, assign 32 one 62 of the set 26 of different loss concealment tools 28 to a portion 22 of the audio signal 12 affected by loss based on
- 30 the first and second measures, and recover the portion 22 of the audio signal using the one loss concealment tool 62 assigned to the portion 22. The derivation of the spectrum might involve, as described, subjecting the scaler factors coded in the data stream to spectral interpolation. Additionally or alternatively, they may be subject to de-emphasis filtering, i.e. they might be multiplied by a de-emphasis filter's transfer function. The
- 35 resulting scale factors may then be subject to spectral centroid measurement. All the other

details described above may then be applied as well. That is, to mention examples which are not meant be exclusively: The set 26 of different loss concealment tools may comprise a first loss concealment tool for audio signal recovery of monophonic portions, and a second loss concealment tool for audio signal recovery of polyphonic portions, and the
5 audio decoder may be configured to, in assigning the one of the set of different loss concealment tools to the portion of the audio signal based on the first and second measures, assign the first loss concealment tool to the portion the more likely the lower the spectral position of the spectral centroid is and the higher the temporal predictability is, and assign the second loss concealment tool to the portion the more likely the higher the
10 spectral position of the spectral centroid is and the lower the temporal predictability is. Additionally or alternatively, the audio decoder may be configured to, in assigning one of the set of different loss concealment tools to a portion 22 of the audio signal affected by loss based on the first and second measures, perform a summation over the first and second measures 42, 52 so as to obtain a scalar sum value and subjecting the scalar sum
15 value to thresholding.

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
20 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 executed by (or using) a hardware apparatus, like for example, a microprocessor, a programmable computer or an electronic circuit. In some embodiments, one or more of the most important method steps may be executed by such an apparatus.

25 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
30 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
35 electronically readable control signals, which are capable of cooperating with a

programmable computer system, such that one of the methods described herein is performed.

5 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.

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

15 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.

20 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.

25 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.

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

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

35 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, for example, comprise a file server for transferring the computer program to the receiver.

5

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,

10

the 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.

15

The apparatus described herein, or any components of the apparatus described herein, may be implemented at least partially in hardware and/or in software.

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 methods described herein, or any components of the apparatus described herein, may be performed at least partially by hardware and/or by software.

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.

25

30

[1] 3GPP TS 26.445; Codec for Enhanced Voice Services (EVS); Detailed algorithmic description.

35

[2] ITU-T G.718 : Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s.

- [3] ITU-T G.711 Appendix I: A high quality low-complexity algorithm for packet loss concealment with G.711.
- [4] 3GPP TS 26.447; Codec for Enhanced Voice Services (EVS); Error concealment of lost packets.
- 5 [5] Method and device for efficient frame erasure concealment in speech codecs; WO2007073604 (A1) — 2007-07-05
- [6] Selecting a Packet Loss Concealment Procedure; EP3111624 A1 — 2017-01-04

Claims

1. Audio decoder for decoding an audio signal from a data stream, the audio decoder
5 comprising a set of different loss concealment tools and configured to

determine a first measure measuring a spectral position of a spectral centroid of a spectrum
of the audio signal,

10 determine a second measure measuring a temporal predictability of the audio signal,

assign one of the set of different loss concealment tools to a portion of the audio signal
affected by loss based on the first and second measures, and

15 recover the portion of the audio signal using the one loss concealment tool assigned to the
portion,

wherein the set of different loss concealment tools comprises:

20 a first loss concealment tool for audio signal recovery of monophonic portions, and

a second loss concealment tool for audio signal recovery of polyphonic portions, and

wherein the audio decoder is configured to, in assigning the one of the set of different loss
25 concealment tools to the portion of the audio signal based on the first and second measures,
assign the first loss concealment tool to the portion the more likely the lower the spectral
position of the spectral centroid is and the higher the temporal predictability is, and assign
the second loss concealment tool to the portion the more likely the higher the spectral
position of the spectral centroid is and the lower the temporal predictability is.
30
2. Audio decoder of claim 1, wherein

the first loss concealment tool is configured to recover the audio signal by audio signal
synthesis using a periodic signal of a periodicity which depends on a pitch value derived
35 from the data stream, and

the second loss concealment tool is configured to cover the audio signal by detecting tonal spectral components of the audio signal, performing phase detection at the tonal spectral components and audio signal synthesis by combining the signals of periodicities which depend on the tonal spectral components with adjusting a mutual phase shift between the signals depending on the phase detection, and

wherein the set of different loss concealment tools further comprises one or more of a third loss concealment tool for audio signal recovery by use of frame repetition, with or without replicate modification.

10

3. Audio decoder of claim 1 or 2, wherein

the first loss concealment tool for audio signal recovery uses tonal time domain packet loss concealment, and

15

the second loss concealment tool for audio signal recovery uses tonal frequency domain packet loss concealment.

4. Audio decoder of any one of the preceding claims, wherein

20

the first loss concealment tool for audio signal recovery is based on audio signal synthesis using a periodic signal of a periodicity which depends on a pitch value derived from the data stream, and

25 the second loss concealment tool for audio signal recovery is based on detecting tonal spectral components of the audio signal, performing phase detection at the tonal spectral components and audio signal synthesis by combining the signals of periodicities which depend on the tonal spectral components with adjusting a mutual phase shift between the signals depending on the phase detection.

30

5. Audio decoder of claim 4, wherein the audio decoder is configured to derive the pitch value from the data stream by using a most recent pitch parameter conveyed in the data stream as the pitch value.

35 6. Audio decoder of claim 4 or 5, wherein the audio decoder is configured to

perform the detection of the tonal spectral components by identifying co-located peaks in one or more consecutive spectra derived from a most recent non-lost portion of the data stream.

5 7. Audio decoder of claim 6, wherein the data stream has the most recent non-lost portion of the data stream encoded thereinto in spectral domain.

8. Audio decoder of any one of the preceding claims, configured to

10 perform loss detection to detect portions affected by loss,

perform the determination of the first and second measures for the portion responsive to the loss detection detecting the portion affected by loss, and refrain from performing the determination for portions not affected by loss.

15

9. Audio decoder of any one of the preceding claims, further configured to

determine a tonality indicator indicative of a tonality of the audio signal,

20 assign one of first and second subsets of one or more loss concealment tools out of the set of different loss concealment tools, which are mutually disjoint, to the portion of the audio signal based on the tonality indicator, and

perform the assignment of the one of the set of different loss concealment tools to the
25 portion of the audio signal based on the first and second measures by assigning the of the set of different loss concealment tools to the portion of the audio signal based on the first and second measures out of the first subset of one or more loss concealment tools if the first subset is assigned to the portion with performing the recovery of the portion of the audio signal using the one loss concealment tool assigned to the portion, and perform the recovery
30 of the portion of the audio signal using one out of the second subset of the loss concealment tools if the second subset of the loss concealment tools is assigned to the portion.

10. Audio decoder of claim 9, wherein the audio decoder is configured to

35 use a parameter conveyed in the data stream as the third measure.

11. Audio decoder of claim 9 or 10, wherein the audio decoder is configured to
use a presence or non-presence of a pitch parameter in a most recent non-lost frame of the
data stream as the third measure.
- 5
12. Audio decoder of any of claims 9 to 11, wherein
the first loss concealment tool is configured to recover the audio signal by audio signal
synthesis using a periodic signal of a periodicity which depends on a pitch value derived
10 from the data stream, and
the second loss concealment tool is configured to recover the audio signal by detecting
tonal spectral components of the audio signal, performing phase detection at the tonal
spectral components and audio signal synthesis by combining the signals of periodicities
15 which depend on the tonal spectral components with adjusting a mutual phase shift between
the signals depending on the phase detection, and
wherein the set of different loss concealment tools further comprises a third loss
concealment tool for audio signal recovery by use of frame repetition, with or without
20 replicate modification, and
wherein the third loss concealment tool is included in the second subset and the second
and the first and second loss concealment tools are included in the first subset.
- 25 13. Audio decoder of any of the preceding claims, configured to determine the first
measure by a weighted sum of spectral component location values, each weighted using
the spectrum of the audio signal at the respective spectral component location value.
14. Audio decoder of any of the preceding claims, configured to determine the second
30 measure by a correlation measure measuring a self-similarity of the audio signal.
15. Audio decoder of any of the preceding claims, configured to determine the second
measure by deriving a pitch from the audio signal and determining the second measure as
a correlation measure measuring an autocorrelation of the audio signal at a temporal shift
35 which depends on the pitch.

16. Audio decoder for decoding an audio signal from a data stream, the audio decoder comprising a set of different loss concealment tools and configured to:
- determine a first measure measuring a spectral position of a spectral centroid of a spectrum of the audio signal,
 - 5 determine a second measure measuring a temporal predictability of the audio signal,
 - assign one of the set of different loss concealment tools to a portion of the audio signal affected by loss based on the first and second measures, and
 - recover the portion of the audio signal using the one loss concealment tool assigned to the portion,
- 10 wherein the audio decoder is configured to, in assigning one of the set of different loss concealment tools to a portion of the audio signal affected by loss based on the first and second measures, perform a summation over the first and second measures so as to obtain a scalar sum value and subjecting the scalar sum value to thresholding.
- 15
17. Audio decoder of any of the preceding claims, configured to determine the first measure by deriving the spectrum from scale factors in a most recent non-lost portion of the data stream.
- 20
18. Audio decoder of any of the preceding claims, configured to determine the first measure by deriving the spectrum from scale factors in a most recent non-lost portion of the data stream and subjecting the scale factors coded in the data stream to spectral interpolation.
- 25
19. Audio decoder of claim 17 or 18, configured to subject the scale factors to a de-emphasis filtering by multiplication with a de-emphasis filter's transfer function.
20. Method for performing loss concealment in audio decoding an audio signal from a data stream, the method comprising:
- 30
- determining a first measure measuring a spectral position of a spectral centroid of a spectrum of the audio signal,
 - determining a second measure measuring a temporal predictability of the audio signal,
- 35

assigning one of a set of different loss concealment tools to a portion of the audio signal affected by loss based on the first and second measures, and

recovering the portion of the audio signal using the one loss concealment tool assigned to the portion,

wherein the set of different loss concealment tools comprises:

a first loss concealment tool for audio signal recovery of monophonic portions, and
a second loss concealment tool for audio signal recovery of polyphonic portions, and

wherein the assigning the one of the set of different loss concealment tools to the portion of the audio signal based on the first and second measures comprises assigning the first loss concealment tool to the portion the more likely the lower the spectral position of the spectral centroid is and the higher the temporal predictability is, and assigning the second loss concealment tool to the portion the more likely the higher the spectral position of the spectral centroid is and the lower the temporal predictability is.

21. Method for performing loss concealment in audio decoding an audio signal from a data stream, the method comprising:

determining a first measure measuring a spectral position of a spectral centroid of a spectrum of the audio signal,

determining a second measure measuring a temporal predictability of the audio signal,

assigning one of a set of different loss concealment tools to a portion of the audio signal affected by loss based on the first and second measures, and

recovering the portion of the audio signal using the one loss concealment tool assigned to the portion,

wherein the assigning one of the set of different loss concealment tools to a portion of the audio signal affected by loss based on the first and second measures comprises performing a summation over the first and second measures so as to obtain a scalar sum value and subjecting the scalar sum value to thresholding.

22. Computer program with a program code for executing, when running on a processor, a method according to claim 20 or 21.