METHOD AND APPARATUS FOR CONTROLLING AUDIO FRAME LOSS CONCEALMENT

TECHNICAL FIELD

The application relates to methods and apparatuses for controlling a concealment method for a lost audio frame of a received audio signal.

BACKGROUND

Conventional audio communication systems transmit speech and audio signals in frames, meaning that the sending side first arranges the signal in short segments or frames of e.g. 20-40 ms which subsequently are encoded and transmitted as a logical unit in e.g. a transmission packet. The receiver decodes each of these units and reconstructs the corresponding signal frames, which in turn are finally output as continuous sequence of reconstructed signal samples. Prior to encoding there is usually an analog to digital (A/D) conversion step that converts the analog speech or audio signal from a microphone into a sequence of audio samples. Conversely, at the receiving end, there is typically a final D/A conversion step that converts the sequence of reconstructed digital signal samples into a time continuous analog signal for loudspeaker playback.

However, such transmission system for speech and audio signals may suffer from transmission errors, which could lead to a situation in which one or several of the transmitted frames are not available at the receiver for reconstruction. In that case, the decoder has to generate a substitution signal for each of the erased, i.e. unavailable frames. This is done in the so-called frame loss or error concealment unit of the receiver-side signal decoder. The purpose of the frame loss concealment is to make the frame loss as inaudible as possible and hence to mitigate the impact of the frame loss on the reconstructed signal quality as much as possible.

Conventional frame loss concealment methods may depend on the structure or architecture of the codec, e.g. by applying a form of repetition of previously received codec parameters. Such parameter repetition techniques are clearly dependent on the specific parameters of the used codec and hence not easily applicable for other codecs with a different structure. Current frame loss concealment methods may e.g. apply the concept of freezing and extrapolating parameters of a previously received frame in order to generate a substitution frame for the lost frame.

These state of the art frame loss concealment methods incorporate some burst loss handling schemes. In general, after a number of frame losses in a row the synthesized signal is attenuated until it is completely muted after long bursts of errors. In addition the coding parameters that are essentially repeated and extrapolated are modified such that the attenuation is accomplished and that spectral peaks are flattened out.

Current state-of-the-art frame loss concealment techniques typically apply the concept of freezing and extrapolating parameters of a previously received frame in order to generate a substitution frame for the lost frame. Many parametric speech codecs such as linear predictive codecs like AMR or AMR-WB typically freeze the earlier received parameters or use some extrapolation thereof and use the decoder with them. In essence, the principle is to have a given model for coding/decoding and to apply the same model with frozen or extrapolated parameters. The frame loss concealment techniques of the AMR and AMR-WB can be regarded as representative. They are specified in detail in the corresponding standards specifications.

Many codecs out of the class of audio codecs apply for coding frequency domain techniques. This means that after some frequency domain transform a coding model is applied on spectral parameters. The decoder reconstructs the signal spectrum from the received parameters and finally transforms the spectrum back to a time signal. Typically, the time signal is reconstructed frame by frame. Such frames are combined by overlap-add techniques to the final reconstructed signal. Even in that case of audio codecs, state-of-the-art error concealment typically applies the same or at least a similar decoding model for lost frames. The frequency domain parameters from a previously received frame are frozen or suitably extrapolated and then used in the frequency-to-time domain conversion. Examples for such techniques are provided with the 3GPP audio codecs according to 3GPP standards.

SUMMARY

Current state-of-the-art solutions for frame loss concealment typically suffer from quality impairments. The main problem is that the parameter freezing and extrapolation technique and re-application of the same decoder model even for lost frames does not always guarantee a smooth and faithful signal evolution from the previously decoded signal frames to the lost frame. This leads typically to audible signal discontinuities with corresponding quality impact.

New schemes for frame loss concealment for speech and audio transmission systems are described. The new schemes improve the quality in case of frame loss over the quality achievable with prior-art frame loss concealment techniques.

The objective of the present embodiments is to control a frame loss concealment scheme that preferably is of the type of the related new methods described such that the best possible sound quality of the reconstructed signal is achieved. The embodiments aim at optimizing this reconstruction quality both with respect to the properties of the signal and of the temporal distribution of the frame losses. Particularly problematic for the frame loss concealment to provide good quality are cases when the audio signal has strongly varying properties such as energy onsets or offsets or if it is spectrally very fluctuating. In that case the described concealment methods may repeat the onset, offset or spectral fluctuation leading to large deviations from the original signal and corresponding quality loss.

Another problematic case is if bursts of frame losses occur in a row. Conceptually, the scheme for frame loss concealment according to the methods described can cope with such cases, though it turns out that annoying tonal artifacts may still occur. It is another objective of the present embodiments to mitigate such artifacts to the highest possible degree.

According to a first aspect, a method for a decoder of concealing a lost audio frame comprises detecting in a property of the previously received and reconstructed audio signal, or in a statistical property of observed frame losses, a condition for which the substitution of a lost frame provides relatively reduced quality. In case such a condition is detected, modifying the concealment method by selectively adjusting a phase or a spectrum magnitude of a substitution frame spectrum.

According to a second aspect, a decoder is configured to implement a concealment of a lost audio frame, and comprises a controller configured to detect in a property of the previously received and reconstructed audio signal, or in a statistical property of observed frame losses, a condition for which the substitution of a lost frame provides relatively reduced quality. In case such a condition is detected, the controller is configured to modify the concealment method by selectively adjusting a phase or a spectrum magnitude of a substitution frame spectrum.

The decoder can be implemented in a device, such as e.g. a mobile phone.

According to a third aspect, a receiver comprises a decoder according to the second aspect described above.

According to a fourth aspect, a computer program is defined for concealing a lost audio frame, and the computer program comprises instructions which when run by a processor causes the processor to conceal a lost audio frame, in agreement with the first aspect described above.

According to a fifth aspect, a computer program product comprises a computer readable medium storing a computer program according to the above-described fourth aspect.

An advantage with an embodiment addresses the control of adaptations frame loss concealment methods allowing mitigating the audible impact of frame loss in the transmission of coded speech and audio signals even further over the quality achieved with only the described concealment methods. The general benefit of the embodiments is to provide a smooth and faithful evolution of the reconstructed signal even for lost frames. The audible impact of frame losses is greatly reduced in comparison to using state-of-the-art techniques.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of example embodiments of the present invention, reference is now made to the following description taken in connection with the accompanying drawings in which:

Figure 1 shows a rectangular window function.

Figure 2 shows a combination of the Hamming window with the rectangular window.

Figure 3 shows an example of a magnitude spectrum of a window function.

Figure 4 illustrates a line spectrum of an exemplary sinusoidal signal with the frequency f_k.

Figure 5 shows a spectrum of a windowed sinusoidal signal with the frequency f_k.

Figure 6 illustrates bars corresponding to the magnitude of grid points of a DFT, based on an analysis frame.

Figure 7 illustrates a parabola fitting through DFT grid points P1, P2 and P3.

Figure 8 illustrates a fitting of a main lobe of a window spectrum.

Figure 9 illustrates a fitting of main lobe approximation function P through DFT grid points P1 and P2.

Figure 10 is a flow chart illustrating an example method according to embodiments of the invention for controlling a concealment method for a lost audio frame of a received audio signal.

Figure 11 is a flow chart illustrating another example method according to embodiments of the invention for controlling a concealment method for a lost audio frame of a received audio signal.

Figure 12 illustrates another example embodiment of the invention.

Figure 13 shows an example of an apparatus according to an embodiment of the invention.
Figure 14 shows another example of an apparatus according to an embodiment of the invention.
Figure 15 shows another example of an apparatus according to an embodiment of the invention.

DETAILED DESCRIPTION

The new controlling scheme for the new frame loss concealment techniques described involve the following steps as shown in Figure 10. It should be noted that the method can be implemented in a controller in a decoder.

1. Detect conditions in the properties of the previously received and reconstructed audio signal or in the statistical properties of the observed frame losses for which the substitution of a lost frame according to the described methods provides relatively reduced quality, **101**.

2. In case such a condition is detected in step 1, modify the element of the methods according to which the substitution frame spectrum is calculated by $Z(m) = Y(m) \cdot e^{j\theta_k}$ by selectively adjusting the phases or the spectrum magnitudes, **102**.

Sinusoidal analysis

A first step of the frame loss concealment technique to which the new controlling technique may be applied involves a sinusoidal analysis of a part of the previously received signal. The purpose of this sinusoidal analysis is to find the frequencies of the main sinusoids of that signal, and the underlying assumption is that the signal is composed of a limited number of individual sinusoids, i.e. that it is a multi-sine signal of the following type:

$$s(n) = \sum_{k=1}^{K} a_k \cdot \cos(2\pi \frac{f_k}{f_s} \cdot n + \varphi_k).$$

In this equation K is the number of sinusoids that the signal is assumed to consist of. For each of the sinusoids with index $k = 1 \ldots K$, a_k is the amplitude, f_k is the frequency, and φ_k is the phase. The sampling frequency is denominated by f_s and the time index of the time discrete signal samples $s(n)$ by n.

It is of main importance to find as exact frequencies of the sinusoids as possible. While an ideal sinusoidal signal would have a line spectrum with line frequencies f_k, finding their true

values would in principle require infinite measurement time. Hence, it is in practice difficult to find these frequencies since they can only be estimated based on a short measurement period, which corresponds to the signal segment used for the sinusoidal analysis described herein; this signal segment is hereinafter referred to as an analysis frame. Another difficulty is that the signal may in practice be time-variant, meaning that the parameters of the above equation vary over time. Hence, on the one hand it is desirable to use a long analysis frame making the measurement more accurate; on the other hand a short measurement period would be needed in order to better cope with possible signal variations. A good trade-off is to use an analysis frame length in the order of e.g. 20-40 ms.

A preferred possibility for identifying the frequencies of the sinusoids f_k is to make a frequency domain analysis of the analysis frame. To this end the analysis frame is transformed into the frequency domain, e.g. by means of DFT or DCT or similar frequency domain transforms. In case a DFT of the analysis frame is used, the spectrum is given by:

$$X(m) = DFT(w(n) \cdot x(n)) = \sum_{n=0}^{L-1} e^{-j\frac{2\pi}{L}mn} \cdot w(n) \cdot x(n) .$$

In this equation w(n) denotes the window function with which the analysis frame of length L is extracted and weighted. Typical window functions are e.g. rectangular windows that are equal to 1 for n $\in$ [0...L-1] and otherwise 0 as shown in Figure 1. It is assumed here that the time indexes of the previously received audio signal are set such that the analysis frame is referenced by the time indexes n=0...L-1. Other window functions that may be more suitable for spectral analysis are, e.g., Hamming window, Hanning window, Kaiser window or Blackman window. A window function that is found to be particular useful is a combination of the Hamming window with the rectangular window. This window has a rising edge shape like the left half of a Hamming window of length $L1$ and a falling edge shape like the right half of a Hamming window of length $L1$ and between the rising and falling edges the window is equal to 1 for the length of L-$L1$, as shown in Figure 2.

The peaks of the magnitude spectrum of the windowed analysis frame $|X(m)|$ constitute an approximation of the required sinusoidal frequencies f_k. The accuracy of this approximation is however limited by the frequency spacing of the DFT. With the DFT with block length L the accuracy is limited to $\frac{f_s}{2L}$.

Experiments show that this level of accuracy may be too low in the scope of the methods described herein. Improved accuracy can be obtained based on the results of the following consideration:

The spectrum of the windowed analysis frame is given by the convolution of the spectrum of the window function with the line spectrum of the sinusoidal model signal $S(\Omega)$, subsequently sampled at the grid points of the DFT:

$$X(m) = \int_{2\pi} \delta(\Omega - m \cdot \tfrac{2\pi}{L}) \cdot \big(W(\Omega) * S(\Omega)\big) \cdot d\Omega .$$

By using the spectrum expression of the sinusoidal model signal, this can be written as

$$X(m) = \tfrac{1}{2} \int_{2\pi} \delta(\Omega - m \cdot \tfrac{2\pi}{L}) \cdot \sum_{k=1}^{K} a_k \cdot \left((W(\Omega + 2\pi \frac{f_k}{f_s}) \cdot e^{-j\varphi_k} + W(\Omega - 2\pi \frac{f_k}{f_s}) \cdot e^{j\varphi_k} \right) \cdot d\Omega .$$

Hence, the sampled spectrum is given by

$$X(m) = \tfrac{1}{2} \sum_{k=1}^{K} a_k \cdot \left((W(2\pi(\frac{m}{L} + \frac{f_k}{f_s})) \cdot e^{-j\varphi_k} + W(2\pi(\frac{m}{L} - \frac{f_k}{f_s})) \cdot e^{j\varphi_k} \right), \text{ with}$$

m=0…L-1.

Based on this consideration it is assumed that the observed peaks in the magnitude spectrum of the analysis frame stem from a windowed sinusoidal signal with K sinusoids where the true sinusoid frequencies are found in the vicinity of the peaks.
Let m_k be the DFT index (grid point) of the observed k^{th} peak, then the corresponding frequency is $\hat{f}_k = \frac{m_k}{L} \cdot f_s$ which can be regarded an approximation of the true sinusoidal frequency f_k. The true sinusoid frequency f_k can be assumed to lie within the interval

$$\left[\left(m_k - \tfrac{1}{2}\right) \cdot \frac{f_s}{L}, \left(m_k + \tfrac{1}{2}\right) \cdot \frac{f_s}{L} \right].$$

For clarity it is noted that the convolution of the spectrum of the window function with the spectrum of the line spectrum of the sinusoidal model signal can be understood as a superposition of frequency-shifted versions of the window function spectrum, whereby the shift frequencies are the frequencies of the sinusoids. This superposition is then sampled at the DFT grid points. These steps are illustrated by the following figures. Figure 3 displays an example of the magnitude spectrum of a window function. Figure 4 shows the magnitude spectrum (line spectrum) of an example sinusoidal signal with a single sinusoid of frequency. Figure 5 shows the magnitude spectrum of the windowed sinusoidal signal that

replicates and superposes the frequency-shifted window spectra at the frequencies of the sinusoid. The bars in Figure 6 correspond to the magnitude of the grid points of the DFT of the windowed sinusoid that are obtained by calculating the DFT of the analysis frame. It should be noted that all spectra are periodic with the normalized frequency parameter Ω where $\Omega = 2\pi$ that corresponds to the sampling frequency f_s.

The previous discussion and the illustration of figure 6 suggest that a better approximation of the true sinusoidal frequencies can only be found through increasing the resolution of the search over the frequency resolution of the used frequency domain transform.

One preferred way to find better approximations of the frequencies f_k of the sinusoids is to apply parabolic interpolation. One such approach is to fit parabolas through the grid points of the DFT magnitude spectrum that surround the peaks and to calculate the respective frequencies belonging to the parabola maxima. A suitable choice for the order of the parabolas is 2. In detail the following procedure can be applied:

1. Identify the peaks of the DFT of the windowed analysis frame. The peak search will deliver the number of peaks K and the corresponding DFT indexes of the peaks. The peak search can typically be made on the DFT magnitude spectrum or the logarithmic DFT magnitude spectrum.

2. For each peak k (with $k = 1 \ldots K$) with corresponding DFT index m_k fit a parabola through the three points $\{P1; P2; P3\} = \{(m_k - 1, \log(|X(m_k - 1)|); (m_k, \log(|X(m_k)|); (m_k + 1, \log(|X(m_k + 1)|)\}$. This results in parabola coefficients $b_k(0)$, $b_k(1)$, $b_k(2)$ of the parabola defined by

$$p_k(q) = \sum_{i=0}^{2} b_k(i) \cdot q^i .$$

This parabola fitting is illustrated in Figure 7.

3. For each of the K parabolas calculate the interpolated frequency index $\hat{m}_k$ corresponding to the value of q for which the parabola has its maximum. Use $f_k = \hat{m}_k \cdot f_s / L$ as approximation for the sinusoid frequency f_k

The described approach provides good results but may have some limitations since the parabolas do not approximate the shape of the main lobe of the magnitude spectrum $|W(\Omega)|$ of the window function. An alternative scheme doing this is an enhanced frequency estimation using a main lobe approximation, described as follows. The main idea of this alternative is to fit a function $P(q)$, which approximates the main lobe of $|W(\frac{2\pi}{L} \cdot q)|$,

through the grid points of the DFT magnitude spectrum that surround the peaks and to calculate the respective frequencies belonging to the function maxima. The function $P(q)$ could be identical to the frequency-shifted magnitude spectrum $|W(\frac{2\pi}{L}\cdot(q-\hat{q}))|$ of the window function. For numerical simplicity it should however rather for instance be a polynomial which allows for straightforward calculation of the function maximum. The following detailed procedure can be applied:

1. Identify the peaks of the DFT of the windowed analysis frame. The peak search will deliver the number of peaks K and the corresponding DFT indexes of the peaks. The peak search can typically be made on the DFT magnitude spectrum or the logarithmic DFT magnitude spectrum.

2. Derive the function $P(q)$ that approximates the magnitude spectrum $|W(\frac{2\pi}{L}\cdot q)|$ of the window function or of the logarithmic magnitude spectrum $\log|W(\frac{2\pi}{L}\cdot q)|$ for a given interval (q_1,q_2). The choice of the approximation function approximating the window spectrum main lobe is illustrated by Figure 8.

3. For each peak k (with k = 1…K) with corresponding DFT index m_k fit the frequency-shifted function $P(q-\hat{q}_k)$ through the two DFT grid points that surround the expected true peak of the continuous spectrum of the windowed sinusoidal signal. Hence, if $|X(m_k - 1)|$ is larger than $|X(m_k+1)|$ fit $P(q-\hat{q}_k)$ through the points $\{P_1; P_2\} = \{(m_k-1, \log(|X(m_k-1)|); (m_k, \log(|X(m_k)|)\}$ and otherwise through the points $\{P_1; P_2\} = \{(m_k, \log(|X(m_k)|); (m_k+1, \log(|X(m_k+1)|)\}$. $P(q)$ can for simplicity be chosen to be a polynomial either of order 2 or 4. This renders the approximation in step 2 a simple linear regression calculation and the calculation of $\hat{q}_k$ straightforward. The interval (q_1,q_2) can be chosen to be fixed and identical for all peaks, e.g. $(q_1,q_2) = (-1,1)$, or adaptive. In the adaptive approach the interval can be chosen such that the function $P(q-\hat{q}_k)$ fits the main lobe of the window function spectrum in the range of the relevant DFT grid points $\{P_1; P_2\}$. The fitting process is visualized in Figure 9.

4. For each of the K frequency shift parameters $\hat{q}_k$ for which the continuous spectrum of the windowed sinusoidal signal is expected to have its peak calculate $\hat{f}_k = \hat{q}_k \cdot f_s/L$ as approximation for the sinusoid frequency f_k.

There are many cases where the transmitted signal is harmonic meaning that the signal consists of sine waves which frequencies are integer multiples of some fundamental frequency f_0. This is the case when the signal is very periodic like for instance for voiced speech or the sustained tones of some musical instrument. This means that the frequencies of the sinusoidal model of the embodiments are not independent but rather have a harmonic relationship and stem from the same fundamental frequency. Taking this harmonic property into account can consequently improve the analysis of the sinusoidal component frequencies substantially.

One enhancement possibility is outlined as follows:

1. Check whether the signal is harmonic. This can for instance be done by evaluating the periodicity of signal prior to the frame loss. One straightforward method is to perform an autocorrelation analysis of the signal. The maximum of such autocorrelation function for some time lag $\tau > 0$ can be used as an indicator. If the value of this maximum exceeds a given threshold, the signal can be regarded harmonic. The corresponding time lag τ then corresponds to the period of the signal which is related to the fundamental frequency through $f_0 = \frac{f_s}{\tau}$.

Many linear predictive speech coding methods apply so-called open or closed-loop pitch prediction or CELP coding using adaptive codebooks. The pitch gain and the associated pitch lag parameters derived by such coding methods are also useful indicators if the signal is harmonic and, respectively, for the time lag.

A further method for obtaining f_0 is described below.

2. For each harmonic index j within the integer range $1...J_{max}$ check whether there is a peak in the (logarithmic) DFT magnitude spectrum of the analysis frame within the vicinity of the harmonic frequency $f_j = j \cdot f_0$. The vicinity of f_j may be defined as the delta range around f_j where delta corresponds to the frequency resolution of the DFT $\frac{f_s}{L}$, i.e. the interval $\left[j \cdot f_0 - \frac{f_s}{2 \cdot L}, j \cdot f_0 + \frac{f_s}{2 \cdot L} \right]$.

In case such a peak with corresponding estimated sinusoidal frequency f_k is present, supersede f_k by $f_k = j \cdot f_0$.

For the two-step procedure given above there is also the possibility to make the check whether the signal is harmonic and the derivation of the fundamental frequency implicitly

and possibly in an iterative fashion without necessarily using indicators from some separate method. An example for such a technique is given as follows:

For each $f_{0,p}$ out of a set of candidate values $\{f_{0,1} \ldots f_{0,P}\}$ apply the procedure step 2, though without superseding f_k but with counting how many DFT peaks are present within the vicinity around the harmonic frequencies, i.e. the integer multiples of $f_{0,p}$. Identify the fundamental frequency $f_{0,p\max}$ for which the largest number of peaks at or around the harmonic frequencies is obtained. If this largest number of peaks exceeds a given threshold, then the signal is assumed to be harmonic. In that case $f_{0,p\max}$ can be assumed to be the fundamental frequency with which step 2 is then executed leading to enhanced sinusoidal frequencies f_k. A more preferable alternative is however first to optimize the fundamental frequency f_0 based on the peak frequencies f_k that have been found to coincide with harmonic frequencies. Assume a set of M harmonics, i.e. integer multiples $\{n_1 \ldots n_M\}$ of some fundamental frequency that have been found to coincide with some set of M spectral peaks at frequencies $f_{k(m)}$, $m = 1 \ldots M$, then the underlying (optimized) fundamental frequency $f_{0,opt}$ can be calculated to minimize the error between the harmonic frequencies and the spectral peak frequencies. If the error to be minimized is the mean square error $E_2 = \sum_{m=1}^{M} (n_m \cdot f_0 - \hat{f}_{k(m)})^2$, then the optimal fundamental frequency is calculated as

$$f_{0,opt} = \frac{\sum_{m=1}^{M} n_m \cdot \hat{f}_{k(m)}}{\sum_{m=1}^{M} n_m^2}.$$

The initial set of candidate values $\{f_{0,1} \ldots f_{0,P}\}$ can be obtained from the frequencies of the DFT peaks or the estimated sinusoidal frequencies f_k.

A further possibility to improve the accuracy of the estimated sinusoidal frequencies f_k is to consider their temporal evolution. To that end, the estimates of the sinusoidal frequencies from a multiple of analysis frames can be combined for instance by means of averaging or prediction. Prior to averaging or prediction a peak tracking can be applied that connects the estimated spectral peaks to the respective same underlying sinusoids.

Applying the sinusoidal model

The application of a sinusoidal model in order to perform a frame loss concealment operation described herein may be described as follows.

It is assumed that a given segment of the coded signal cannot be reconstructed by the decoder since the corresponding encoded information is not available. It is further assumed that a part of the signal prior to this segment is available. Let $y(n)$ with $n = 0 \ldots N\text{-}1$ be the unavailable segment for which a substitution frame $z(n)$ has to be generated and $y(n)$ with $n<0$ be the available previously decoded signal. Then, in a first step a prototype frame of the available signal of length L and start index n_{-1} is extracted with a window function $w(n)$ and transformed into frequency domain, e.g. by means of DFT:

$$Y_{-1}(m) = \sum_{n=0}^{L-1} y(n - n_{-1}) \cdot w(n) \cdot e^{-j\frac{2\pi}{L} n m} .$$

The window function can be one of the window functions described above in the sinusoidal analysis. Preferably, in order to save numerical complexity, the frequency domain transformed frame should be identical with the one used during sinusoidal analysis.

In a next step the sinusoidal model assumption is applied. According to that the DFT of the prototype frame can be written as follows:

$$Y_{-1}(m) = \frac{1}{2} \sum_{k=1}^{K} a_k \cdot \left((W(2\pi(\frac{m}{L} + \frac{f_k}{f_s})) \cdot e^{-j\varphi_k} + W(2\pi(\frac{m}{L} - \frac{f_k}{f_s})) \cdot e^{j\varphi_k}) \right) .$$

The next step is to realize that the spectrum of the used window function has only a significant contribution in a frequency range close to zero. As illustrated in Figure 3 the magnitude spectrum of the window function is large for frequencies close to zero and small otherwise (within the normalized frequency range from $-\pi$ to π, corresponding to half the sampling frequency). Hence, as an approximation it is assumed that the window spectrum $W(m)$ is non-zero only for an interval $M = [-m_{min}, m_{max}]$, with m_{min} and m_{max} being small positive numbers. In particular, an approximation of the window function spectrum is used such that for each k the contributions of the shifted window spectra in the above expression are strictly non-overlapping. Hence in the above equation for each frequency index there is always only at maximum the contribution from one summand, i.e. from one shifted window spectrum. This means that the expression above reduces to the following approximate expression:

$$\hat{Y}_{-1}(m) = \frac{a_k}{2} \cdot W\left(2\pi\left(\frac{m}{L} - \frac{f_k}{f_s}\right)\right) \cdot e^{j\varphi_k}$$ for non-negative $m \in M_k$ and for each k.

Herein, M_k denotes the integer interval

$M_k = [\text{round}\left(\frac{f_k}{f_s} \cdot L\right) - m_{min,k}, \text{round}\left(\frac{f_k}{f_s} \cdot L\right) + m_{max,k}]$, where $m_{min,k}$ and $m_{max,k}$ fulfill the above explained constraint such that the intervals are not overlapping. A suitable choice for $m_{min,k}$ and $m_{max,k}$ is to set them to a small integer value δ, e.g. $\delta = 3$. If however the DFT indices related to two neighboring sinusoidal frequencies f_k and f_{k+1} are less than 2δ, then δ is set to $\text{floor}\left(\frac{\text{round}\left(\frac{f_{k+1}}{f_s} \cdot L\right) - \text{round}\left(\frac{f_k}{f_s} \cdot L\right)}{2}\right)$ such that it is ensured that the intervals are not overlapping. The function floor $(\cdot)$ is the closest integer to the function argument that is smaller or equal to it.

The next step according to the embodiment is to apply the sinusoidal model according to the above expression and to evolve its K sinusoids in time. The assumption that the time indices of the erased segment compared to the time indices of the prototype frame differs by n_{-1} samples means that the phases of the sinusoids advance by

$$\theta_k = 2\pi \cdot \frac{f_k}{f_s} n_{-1} \quad .$$

Hence, the DFT spectrum of the evolved sinusoidal model is given by:

$$Y_0(m) = \frac{1}{2}\sum_{k=1}^{K} a_k \cdot \left((W(2\pi(\frac{m}{L} + \frac{f_k}{f_s})) \cdot e^{-j(\varphi_k + \theta_k)} + W(2\pi(\frac{m}{L} - \frac{f_k}{f_s})) \cdot e^{j(\varphi_k + \theta_k)} \right) \quad .$$

Applying again the approximation according to which the shifted window function spectra do no overlap gives:

$Y_0(m) = \frac{a_k}{2} \cdot W\left(2\pi\left(\frac{m}{L} - \frac{f_k}{f_s}\right)\right) \cdot e^{j(\varphi_k + \theta_k)}$ for non-negative $m \in M_k$ and for each k.

Comparing the DFT of the prototype frame $Y_{-1}(m)$ with the DFT of evolved sinusoidal model $Y_0(m)$ by using the approximation, it is found that the magnitude spectrum remains unchanged while the phase is shifted by $\theta_k = 2\pi \cdot \frac{f_k}{f_s} n_{-1}$, for each $m \in M_k$. Hence, the frequency spectrum coefficients of the prototype frame in the vicinity of each sinusoid are shifted proportional to the sinusoidal frequency f_k and the time difference between the lost audio frame and the prototype frame n_{-1}.

Hence, according to the embodiment the substitution frame can be calculated by the following expression:

$z(n) = IDTF\{Z(m)\}$ with $Z(m) = Y(m) \cdot e^{j\theta_k}$ for non-negative $m \in M_k$ and for each k.

A specific embodiment addresses phase randomization for DFT indices not belonging to any interval M_k. As described above, the intervals M_k, $k = 1 \ldots K$ have to be set such that they are strictly non-overlapping which is done using some parameter δ which controls the size of the intervals. It may happen that δ is small in relation to the frequency distance of two neighboring sinusoids. Hence, in that case it happens that there is a gap between two intervals. Consequently, for the corresponding DFT indices m no phase shift according to the above expression $Z(m) = Y(m) \cdot e^{j\theta_k}$ is defined. A suitable choice according to this embodiment is to randomize the phase for these indices, yielding $Z(m) = Y(m) \cdot e^{j2\pi \mathrm{rand}(\cdot)}$, where the function rand(·) returns some random number.

It has been found beneficial for the quality of the reconstructed signals to optimize the size of the intervals M_k. In particular, the intervals should be larger if the signal is very tonal, i.e. when it has clear and distinct spectral peaks. This is the case for instance when the signal is harmonic with a clear periodicity. In other cases where the signal has less pronounced spectral structure with broader spectral maxima, it has been found that using small intervals leads to better quality. This finding leads to a further improvement according to which the interval size is adapted according to the properties of the signal. One realization is to use a tonality or a periodicity detector. If this detector identifies the signal as tonal, the δ-parameter controlling the interval size is set to a relatively large value. Otherwise, the δ-parameter is set to relatively smaller values.

Based on the above, the audio frame loss concealment methods involve the following steps:

1. Analyzing a segment of the available, previously synthesized signal to obtain the constituent sinusoidal frequencies f_k of a sinusoidal model, optionally using an enhanced frequency estimation.

2. Extracting a prototype frame y_{-1} from the available previously synthesized signal and calculate the DFT of that frame.

3. Calculating the phase shift θ_k for each sinusoid k in response to the sinusoidal frequency f_k and the time advance n_{-1} between the prototype frame and the substitution frame. Optionally in this step the size of the interval M may have been adapted in response to the tonality of the audio signal.

4. For each sinusoid k advancing the phase of the prototype frame DFT with θ_k selectively for the DFT indices related to a vicinity around the sinusoid frequency f_k.

5. Calculating the inverse DFT of the spectrum obtained in step 4.

Signal and frame loss property analysis and detection

The methods described above are based on the assumption that the properties of the audio signal do not change significantly during the short time duration from the previously received and reconstructed signal frame and a lost frame. In that case it is a very good choice to retain the magnitude spectrum of the previously reconstructed frame and to evolve the phases of the sinusoidal main components detected in the previously reconstructed signal. There are however cases where this assumption is wrong which are for instance transients with sudden energy changes or sudden spectral changes.

A first embodiment of a transient detector according to the invention can consequently be based on energy variations within the previously reconstructed signal. This method, illustrated in Figure 11, calculates the energy in a left part and a right part of some analysis frame **113**. The analysis frame may be identical to the frame used for sinusoidal analysis described above. A part (either left or right) of the analysis frame may be the first or respectively the last half of the analysis frame or e.g. the first or respectively the last quarter of the analysis frame, **110**. The respective energy calculation is done by summing the squares of the samples in these partial frames:

$$E_{left} = \sum_{n=0}^{N_{part}-1} y^2(n - n_{left}), \text{ and } E_{right} = \sum_{n=0}^{N_{part}-1} y^2(n - n_{right}).$$

Herein $y(n)$ denotes the analysis frame, n_{left} and n_{right} denote the respective start indices of the partial frames that are both of size N_{part}.

Now the left and right partial frame energies are used for the detection of a signal discontinuity. This is done by calculating the ratio

$$R_{l/r} = \frac{E_{left}}{E_{right}}.$$

A discontinuity with sudden energy decrease (offset) can be detected if the ratio $R_{l/r}$ exceeds some threshold (e.g. 10), **115**. Similarly a discontinuity with sudden energy

increase (onset) can be detected if the ratio $R_{l/r}$ is below some other threshold (e.g. 0.1), **117**.

In the context of the above described concealment methods it has been found that the above defined energy ratio may in many cases be a too insensitive indicator. In particular in real signals and especially music there are cases where a tone at some frequency suddenly emerges while some other tone at some other frequency suddenly stops. Analyzing such a signal frame with the above-defined energy ratio would in any case lead to a wrong detection result for at least one of the tones since this indicator is insensitive to different frequencies.

A solution to this problem is described in the following embodiment. The transient detection is now done in the time frequency plane. The analysis frame is again partitioned into a left and a right partial frame, **110**. Though now, these two partial frames are (after suitable windowing with e.g. a Hamming window, **111**) transformed into the frequency domain, e.g. by means of a N_{part} -point DFT, **112**.

$$Y_{left}(m) = DFT\{y(n - n_{left})\}_{N_{part}} \text{ and}$$

$$Y_{right}(m) = DFT\{y(n - n_{right})\}_{N_{part}}, \text{ with } m = 0 \ldots N_{part} - 1.$$

Now the transient detection can be done frequency selectively for each DFT bin with index m. Using the powers of the left and right partial frame magnitude spectra, for each DFT index m a respective energy ratio can be calculated **113** as

$$R_{l/r}(m) = \frac{|Y_{left}(m)|^2}{|Y_{right}(m)|^2}.$$

Experiments show that frequency selective transient detection with DFT bin resolution is relatively imprecise due to statistical fluctuations (estimation errors). It was found that the quality of the operation is rather enhanced when making the frequency selective transient detection on the basis of frequency bands. Let $l_k = [m_{k-1} + 1, \ldots, m_k]$ specify the k^{th} interval, $k = 1 \ldots K$, covering the DFT bins from $m_{k-1} + 1$ to m_k, then these intervals define K frequency bands. The frequency group selective transient detection can now be based on the band-wise ratio between the respective band energies of the left and right partial frames:

$$R_{left,band}(k) = \frac{\sum_{m \in I_k} |V_{left}(m)|^2}{\sum_{m \in I_k} |V_{right}(m)|^2}.$$

It is to be noted that the interval $I_k = [m_{k-1} + 1, ..., m_k]$ corresponds to the frequency band $B_k = [\frac{m_{k-1}+1}{N_{part}} \cdot f_s, ..., \frac{m_k}{N_{part}} \cdot f_s]$, where f_s denotes the audio sampling frequency.

The lowest lower frequency band boundary m_0 can be set to 0 but may also be set to a DFT index corresponding to a larger frequency in order to mitigate estimation errors that grow with lower frequencies. The highest upper frequency band boundary m_K can be set to $\frac{N_{part}}{2}$ but is preferably chosen to correspond to some lower frequency in which a transient still has a significant audible effect.

A suitable choice for these frequency band sizes or widths is either to make them equal size with e.g. a width of several 100 Hz. Another preferred way is to make the frequency band widths following the size of the human auditory critical bands, i.e. to relate them to the frequency resolution of the auditory system. This means approximately to make the frequency band widths equal for frequencies up to 1kHz and to increase them exponentially above 1 kHz. Exponential increase means for instance to double the frequency bandwidth when incrementing the band index k.

As described in the first embodiment of the transient detector that was based on an energy ratio of two partial frames, any of the ratios related to band energies or DFT bin energies of two partial frames are compared to certain thresholds. A respective upper threshold for (frequency selective) offset detection **115** and a respective lower threshold for (frequency selective) onset detection **117** is used.

A further audio signal dependent indicator that is suitable for an adaptation of the frame loss concealment method can be based on the codec parameters transmitted to the decoder. For instance, the codec may be a multi-mode codec like ITU-T G.718. Such codec may use particular codec modes for different signal types and a change of the codec mode in a frame shortly before the frame loss may be regarded as an indicator for a transient.

Another useful indicator for adaptation of the frame loss concealment is a codec parameter related to a voicing property and the transmitted signal. Voicing relates to highly periodic speech that is generated by a periodic glottal excitation of the human vocal tract.

A further preferred indicator is whether the signal content is estimated to be music or speech. Such an indicator can be obtained from a signal classifier that may typically be part of the codec. In case the codec performs such a classification and makes a corresponding classification decision available as a coding parameter to the decoder, this parameter is preferably used as signal content indicator to be used for adapting the frame loss concealment method.

Another indicator that is preferably used for adaptation of the frame loss concealment methods is the burstiness of the frame losses. Burstiness of frame losses means that there occur several frame losses in a row, making it hard for the frame loss concealment method to use valid recently decoded signal portions for its operation. A state-of-the-art indicator is the number n_{burst} of observed frame losses in a row. This counter is incremented with one upon each frame loss and reset to zero upon the reception of a valid frame. This indicator is also used in the context of the present example embodiments of the invention.

Adaptation of the frame loss concealment method

In case the steps carried out above indicate a condition suggesting an adaptation of the frame loss concealment operation the calculation of the spectrum of the substitution frame is modified.

While the original calculation of the substitution frame spectrum is done according to the expression $Z(m) = Y(m) \cdot e^{j\theta}{}_k$, now an adaptation is introduced modifying both magnitude and phase. The magnitude is modified by means of scaling with two factors $\alpha(m)$ and $\beta(m)$ and the phase is modified with an additive phase component $\vartheta(m)$. This leads to the following modified calculation of the substitution frame:

$$Z(m) = \alpha(m) \cdot \beta(m) \cdot Y(m) \cdot e^{j(\theta_k + \vartheta(m))}.$$

It is to be noted that the original (non-adapted) frame-loss concealment methods is used if $\alpha(m) = 1, \beta(m) = 1, \text{and } \vartheta(m) = 0$. These respective values are hence the default.

The general objective with introducing magnitude adaptations is to avoid audible artifacts of the frame loss concealment method. Such artifacts may be musical or tonal sounds or strange sounds arising from repetitions of transient sounds. Such artifacts would in turn lead to quality degradations, which avoidance is the objective of the described adaptations. A suitable way to such adaptations is to modify the magnitude spectrum of the substitution frame to a suitable degree.

Figure 12 illustrates an embodiment of concealment method modification. Magnitude adaptation, **123,** is preferably done if the burst loss counter n_{burst} exceeds some threshold thr_{burst}, e.g. $thr_{burst} = 3$, **121**. In that case a value smaller than 1 is used for the attenuation factor, e.g. $\alpha(m) = 0.1$.

It has however been found that it is beneficial to perform the attenuation with gradually increasing degree. One preferred embodiment which accomplishes this is to define a logarithmic parameter specifying a logarithmic increase in attenuation per frame, att_per_frame. Then, in case the burst counter exceeds the threshold the gradually increasing attenuation factor is calculated by

$$\alpha(m) = 10^{c \cdot att_per_frame \cdot (n_{burst} - thr_{burst})}.$$

Here the constant c is mere a scaling constant allowing to specify the parameter att_per_frame for instance in decibels (dB).

An additional preferred adaptation is done in response to the indicator whether the signal is estimated to be music or speech. For music content in comparison with speech content it is preferable to increase the threshold thr_{burst} and to decrease the attenuation per frame. This is equivalent with performing the adaptation of the frame loss concealment method with a lower degree. The background of this kind of adaptation is that music is generally less sensitive to longer loss bursts than speech. Hence, the original, i.e. the unmodified frame loss concealment method is still preferable for this case, at least for a larger number of frame losses in a row.

A further adaptation of the concealment method with regards to the magnitude attenuation factor is preferably done in case a transient has been detected based on that the indicator $R_{l/r,\,band}(k)$ or alternatively $R_{l/r}(m)$ or $R_{l/r}$ have passed a threshold, **122**. In that case a suitable adaptation action, **125**, is to modify the second magnitude attenuation factor $\beta(m)$ such that the total attenuation is controlled by the product of the two factors $\alpha(m) \cdot \beta(m)$.

$\beta(m)$ is set in response to an indicated transient. In case an offset is detected the factor $\beta(m)$ is preferably be chosen to reflect the energy decrease of the offset. A suitable choice is to set $\beta(m)$ to the detected gain change:

$$\beta(m) = \sqrt{R_{l/r,band}(k)}, \text{ for } m \in I_k, k = 1 \ldots K.$$

In case an onset is detected it is rather found advantageous to limit the energy increase in the substitution frame. In that case the factor can be set to some fixed value of e.g. 1, meaning that there is no attenuation but not any amplification either.

In the above it is to be noted that the magnitude attenuation factor is preferably applied frequency selectively, i.e. with individually calculated factors for each frequency band. In case the band approach is not used, the corresponding magnitude attenuation factors can still be obtained in an analogue way. $\beta(m)$ can then be set individually for each DFT bin in case frequency selective transient detection is used on DFT bin level. Or, in case no frequency selective transient indication is used at all $\beta(m)$ can be globally identical for all m.

A further preferred adaptation of the magnitude attenuation factor is done in conjunction with a modification of the phase by means of the additional phase component $\vartheta(m)$ **127**. In case for a given m such a phase modification is used, the attenuation factor $\beta(m)$ is reduced even further. Preferably, even the degree of phase modification is taken into account. If the phase modification is only moderate, $\beta(m)$ is only scaled down slightly, while if the phase modification is strong, $\beta(m)$ is scaled down to a larger degree.

The general objective with introducing phase adaptations is to avoid too strong tonality or signal periodicity in the generated substitution frames, which in turn would lead to quality degradations. A suitable way to such adaptations is to randomize or dither the phase to a suitable degree.

Such phase dithering is accomplished if the additional phase component $\vartheta(m)$ is set to a random value scaled with some control factor: $\vartheta(m) = \alpha(m) \cdot \mathrm{rand}(\cdot)$.

The random value obtained by the function rand(·) is for instance generated by some pseudo-random number generator. It is here assumed that it provides a random number within the interval $[0, 2\pi]$.

The scaling factor $\alpha(m)$ in the above equation control the degree by which the original phase θ_k is dithered. The following embodiments address the phase adaptation by means of controlling this scaling factor. The control of the scaling factor is done in an analogue way as the control of the magnitude modification factors described above.

According to a first embodiment scaling factor $\alpha(m)$ is adapted in response to the burst loss counter. If the burst loss counter n_{burst} exceeds some threshold thr_{burst}, e.g. $thr_{burst} = 3$, a value larger than 0 is used, e.g. $\alpha(m)$ = 0.2.

It has however been found that it is beneficial to perform the dithering with gradually increasing degree. One preferred embodiment which accomplishes this is to define a parameter specifying an increase in dithering per frame, $dith_increase_per_frame$. Then in case the burst counter exceeds the threshold the gradually increasing dithering control factor is calculated by

$$a(m) = dith_increase_per_frame \cdot (n_{burst} - thr_{burst}).$$

It is to be noted in the above formula that $a(m)$ has to be limited to a maximum value of 1 for which full phase dithering is achieved.

It is to be noted that the burst loss threshold value thr_{burst} used for initiating phase dithering may be the same threshold as the one used for magnitude attenuation. However, better quality can be obtained by setting these thresholds to individually optimal values, which generally means that these thresholds may be different.

An additional preferred adaptation is done in response to the indicator whether the signal is estimated to be music or speech. For music content in comparison with speech content it is preferable to increase the threshold thr_{burst} meaning that phase dithering for music as compared to speech is done only in case of more lost frames in a row. This is equivalent with performing the adaptation of the frame loss concealment method for music with a lower degree. The background of this kind of adaptation is that music is generally less sensitive to longer loss bursts than speech. Hence, the original, i.e. unmodified frame loss concealment method is still preferable for this case, at least for a larger number of frame losses in a row.

A further preferred embodiment is to adapt the phase dithering in response to a detected transient. In that case a stronger degree of phase dithering can be used for the DFT bins m for which a transient is indicated either for that bin, the DFT bins of the corresponding frequency band or of the whole frame.

Part of the schemes described address optimization of the frame loss concealment method for harmonic signals and particularly for voiced speech.

In case the methods using an enhanced frequency estimation as described above are not realized another adaptation possibility for the frame loss concealment method optimizing the quality for voiced speech signals is to switch to some other frame loss concealment method that specifically is designed and optimized for speech rather than for general audio signals containing music and speech. In that case, the indicator that the signal comprises a voiced

speech signal is used to select another speech-optimized frame loss concealment scheme rather than the schemes described above.

The embodiments apply to a controller in a decoder, as illustrated in Figure 13. Figure 13 is a schematic block diagram of a decoder according to the embodiments. The decoder **130** comprises an input unit **132** configured to receive an encoded audio signal. The figure illustrates the frame loss concealment by a logical frame loss concealment-unit **134**, which indicates that the decoder is configured to implement a concealment of a lost audio frame, according to the above-described embodiments. Further the decoder comprises a controller **136** for implementing the embodiments described above. The controller **136** is configured to detect conditions in the properties of the previously received and reconstructed audio signal or in the statistical properties of the observed frame losses for which the substitution of a lost frame according to the described methods provides relatively reduced quality. In case such a condition is detected, the controller **136** is configured to modify the element of the concealment methods according to which the substitution frame spectrum is calculated by $Z(m) = Y(m) \cdot e^{j\theta_k}$ by selectively adjusting the phases or the spectrum magnitudes. The detection can be performed by a detector unit **146** and modifying can be performed by a modifier unit **148** as illustrated in Figure 14.

The decoder with its including units could be implemented in hardware. There are numerous variants of circuitry elements that can be used and combined to achieve the functions of the units of the decoder. Such variants are encompassed by the embodiments. Particular examples of hardware implementation of the decoder is implementation in digital signal processor (DSP) hardware and integrated circuit technology, including both general-purpose electronic circuitry and application-specific circuitry.

The decoder **150** described herein could alternatively be implemented e.g. as illustrated in Figure 15, i.e. by one or more of a processor **154** and adequate software **155** with suitable storage or memory **156** therefore, in order to reconstruct the audio signal, which includes performing audio frame loss concealment according to the embodiments described herein, as shown in Figure 13. The incoming encoded audio signal is received by an input (IN) **152**, to which the processor **154** and the memory **156** are connected. The decoded and reconstructed audio signal obtained from the software is outputted from the output (OUT) **158**.

The technology described above may be used e.g. in a receiver, which can be used in a mobile device (e.g. mobile phone, laptop) or a stationary device, such as a personal computer.

It is to be understood that the choice of interacting units or modules, as well as the naming of the units are only for exemplary purpose, and may be configured in a plurality of alternative ways in order to be able to execute the disclosed process actions.

It should also be noted that the units or modules described in this disclosure are to be regarded as logical entities and not with necessity as separate physical entities. It will be appreciated that the scope of the technology disclosed herein fully encompasses other embodiments which may become obvious to those skilled in the art, and that the scope of this disclosure is accordingly not to be limited.

Reference to an element in the singular is not intended to mean "one and only one" unless explicitly so stated, but rather "one or more." All structural and functional equivalents to the elements of the above-described embodiments that are known to those of ordinary skill in the art are expressly incorporated herein by reference and are intended to be encompassed hereby. Moreover, it is not necessary for a device or method to address each and every problem sought to be solved by the technology disclosed herein, for it to be encompassed hereby.

In the preceding description, for purposes of explanation and not limitation, specific details are set forth such as particular architectures, interfaces, techniques, etc. in order to provide a thorough understanding of the disclosed technology. However, it will be apparent to those skilled in the art that the disclosed technology may be practiced in other embodiments and/or combinations of embodiments that depart from these specific details. That is, those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the disclosed technology. In some instances, detailed descriptions of well-known devices, circuits, and methods are omitted so as not to obscure the description of the disclosed technology with unnecessary detail. All statements herein reciting principles, aspects, and embodiments of the disclosed technology, as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents as well as equivalents developed in the future, e.g. any elements developed that perform the same function, regardless of structure.

Thus, for example, it will be appreciated by those skilled in the art that the figures herein can represent conceptual views of illustrative circuitry or other functional units embodying the principles of the technology, and/or various processes which may be substantially represented in computer readable medium and executed by a computer or processor, even though such computer or processor may not be explicitly shown in the figures.

The functions of the various elements including functional blocks may be provided through the use of hardware such as circuit hardware and/or hardware capable of executing software in the form of coded instructions stored on computer readable medium. Thus, such functions and illustrated functional blocks are to be understood as being either hardware-implemented and/or computer-implemented, and thus machine-implemented.

The embodiments described above are to be understood as a few illustrative examples of the present invention. It will be understood by those skilled in the art that various modifications, combinations and changes may be made to the embodiments without departing from the scope of the present invention. In particular, different part solutions in the different embodiments can be combined in other configurations, where technically possible.

CLAIMS

1. A method for controlling a concealment method for a lost audio frame of a received audio signal, the method comprising:

- detecting (101) in a property of a previously received and reconstructed audio signal, or in a statistical property of observed frame losses, a condition for which a substitution of a lost frame provides a relatively reduced quality; and
- modifying (102) the concealment method by selectively adjusting a phase or a spectrum magnitude of a substitution frame spectrum, when a condition is detected.

2. The method according to claim 1, wherein original calculation of the substitution frame spectrum is performed according to the expression $Z(m) = Y(m) \cdot e^{j\theta_k}$.

3. The method according to claim 1 or 2, wherein the detected condition comprises a transient detection.

4. The method according to claim 3, wherein the transient detection is performed in a frequency domain.

5. The method according to claim 3 or 4, wherein the transient detection comprises:
- partitioning an analysis frame into two partial frames;
- calculating an energy ratio of the two partial frames; and
- comparing the energy ratio to a defined threshold.

6. The method according to claim 5, wherein a first partial frame comprises a left part of the analysis frame and a second partial frame comprises a right part of the analysis frame.

7. The method according to claim 5, wherein the defined threshold comprises an upper threshold for offset detection and a lower threshold for onset detection.

8. The method according to any of the claims 3 to 7, wherein the transient detection is performed frequency selectively on the basis of a frequency band

9. The method according to claim 8, wherein frequency band widths follow the size of the human auditory critical bands.

10. The method according to any of preceding claims, wherein the concealment method is further modified in response to an indicator of a condition for which a substitution of a lost frame provides a relatively reduced quality, the indicator being based on at least one of: a parameter indicating a used codec mode, a parameter related to a voicing property of speech and a signal content indicator indicating whether a signal content is estimated to be music or speech.

11. The method according to claim 10, wherein an alternative frame loss concealment method optimized for speech signals is selected in case the indicator indicates that a signal comprises voiced speech.

12. The method according to claim 1, wherein one statistical property of the observed frame losses for which the substitution of the lost frame provides relatively reduced quality is a burstiness of the frame losses.

13. The method according to claim 12, wherein the spectrum magnitude is adjusted in response to detected burstiness of the frame losses by a gradual increase of a first attenuation factor.

14. The method according to claim 13, wherein a second attenuation factor is set in response to an indicated transient, the total attenuation being controlled by the product of the first and the second attenuation factors.

15. The method according to claim 1, wherein adjusting a phase comprises randomizing or dithering a phase spectrum.

16. The method according to claims 12 and 15, wherein the phase spectrum is adjusted in response to detected burstiness of the frame losses by performing the dithering with gradually increasing degree.

17. An apparatus comprising means for performing a method according to at least one of claims 1 to 16.

18. An apparatus comprising:

a processor (154), and

a memory (156) storing instructions (155) that, when executed by the processor, cause the apparatus to

- detect in a property of a previously received and reconstructed audio signal, or in a statistical property of observed frame losses, a condition for which a substitution of a lost frame provides relatively reduced quality; and
- modify the concealment method, when such a condition is detected, by selectively adjusting a phase or a spectrum magnitude of a substitution frame spectrum.

19. The apparatus according to claim 18, wherein original calculation of the substitution frame spectrum is performed according to the expression $Z(m) = Y(m) \cdot e^{j\theta_k}$.

20. The apparatus according to claim 18 further comprising a transient detector.

21. The apparatus according to claim 20, wherein the transient detector is configured to perform transient detection in the frequency domain.

22. The apparatus according to claim 20 or 21, wherein the transient detector is configured to

- partition an analysis frame into two partial frames;
- calculate an energy ratio of the two partial frames; and
- compare the energy ratio to a defined threshold.

23. The apparatus according to any of claims 20 to 22, wherein the transient detector is configured to perform a frequency selective transient detection on the basis of frequency bands.

24. The apparatus according to any of claims 18 to 23, wherein the apparatus is further configured to modify the concealment method in response to an indicator of a condition for which a substitution of a lost frame provides relatively reduced quality, the indicator being based on at least one of: a parameter indicating a used codec mode, a parameter related to a voicing property of speech and a signal content indicator indicating whether a signal content is estimated to be music or speech.

25. The apparatus according to claim 18, wherein one statistical property of the observed frame losses for which the substitution of the lost frame provides relatively reduced quality is a burstiness of the frame losses.

26. The apparatus according to claim 25, wherein a spectrum magnitude is adjusted in response to a detected burstiness of the frame losses by gradually increasing a first attenuation factor.

27. The apparatus according to claim 26, wherein a second attenuation factor is set in response to an indicated transient, the total attenuation being controlled by the product of the first and the second attenuation factors.

28. The apparatus according to claim 18, wherein adjusting a phase comprises randomizing or dithering a phase spectrum.

29. The apparatus according to claim 17 or 18, wherein the apparatus is a decoder in a mobile device.

30. A computer program (155) comprising computer readable code units which when run on an apparatus causes the apparatus to:

- detect (101) in a property of a previously received and reconstructed audio signal, or in a statistical property of observed frame losses, a condition for which a substitution of a lost frame provides relatively reduced quality; and
- modify (102) the concealment method, when such a condition is detected, by selectively adjusting a phase or a spectrum magnitude of a substitution frame spectrum.

31. A computer program product (156), comprising computer readable medium and a computer program (155) according to claim 30 stored on the computer readable medium.

32. A decoder (130) comprising:

- an input unit (132) configured to receive an encoded audio signal;
- a logical frame loss concealment-unit (134) configured to conceal a lost audio frame;
- a controller (136) configured to detect, in a property of a previously received and reconstructed audio signal, or in a statistical property of observed frame losses, a condition for which the substitution of a lost frame provides relatively reduced quality, and to modify the concealment of a lost audio frame by selectively adjusting a phase or a spectrum magnitude of a substitution frame spectrum, when detecting such a condition.

33. The decoder according to claim 32, wherein the controller (136) comprises a detector unit (146) for performing the detection of a condition in a property of the previously received and reconstructed audio signal, or in the statistical property of the observed frame losses, and a modifier unit (148) for performing the modification of the concealment method.

5

34. An apparatus (130) configured to control a concealment method for a lost audio frame of a received audio signal, the apparatus comprising:
– a detection module (146) for detecting in a property of a previously received and reconstructed audio signal, or in a statistical property of observed frame losses, a condition
10 for which a substitution of a lost frame provides relatively reduced quality; and
– a modification module (148) for modifying the concealment method, when such a condition is detected, by selectively adjusting a phase or a spectrum magnitude of a substitution frame spectrum.