CROSS PRODUCT ENHANCED HARMONIC TRANSPOSITION

TECHNICAL FIELD

The present invention relates to audio coding systems which make use of a harmonic transposition method for high frequency reconstruction (HFR).

BACKGROUND OF THE INVENTION

HFR technologies, such as the Spectral Band Replication (SBR) technology, allow to significantly improve the coding efficiency of traditional perceptual audio codecs. In combination with MPEG-4 Advanced Audio Coding (AAC) it forms a very efficient audio codec, which is already in use within the XM Satellite Radio system and Digital Radio Mondiale. The combination of AAC and SBR is called aacPlus. It is part of the MPEG-4 standard where it is referred to as the High Efficiency AAC Profile. In general, HFR technology can be combined with any perceptual audio codec in a back and forward compatible way, thus offering the possibility to upgrade already established broadcasting systems like the MPEG Layer-2 used in the Eureka DAB system. HFR transposition methods can also be combined with speech codecs to allow wide band speech at ultra low bit rates.

The basic idea behind HRF is the observation that usually a strong correlation between the characteristics of the high frequency range of a signal and the characteristics of the low frequency range of the same signal is present. Thus, a good approximation for the representation of the original input high frequency range of a signal can be achieved by a signal transposition from the low frequency range to the high frequency range.

This concept of transposition was established in WO 98/57436, as a method to recreate a high frequency band from a lower frequency band of an audio signal. A substantial saving in bit-rate can be obtained by using this concept in audio coding and/or speech coding. In the following, reference will be made to audio coding, but it should be noted

that the described methods and systems are equally applicable to speech coding and in unified speech and audio coding (USAC).

In a HFR based audio coding system, a low bandwidth signal is presented to a core waveform coder and the higher frequencies are regenerated at the decoder side using transposition of the low bandwidth signal and additional side information, which is typically encoded at very low bit-rates and which describes the target spectral shape. For low bit-rates, where the bandwidth of the core coded signal is narrow, it becomes increasingly important to recreate a high band, i.e. the high frequency range of the audio signal, with perceptually pleasant characteristics. Two variants of harmonic frequency reconstruction methods are mentioned in the following, one is referred to as harmonic transposition and the other one is referred to as single sideband modulation.

The principle of harmonic transposition defined in WO 98/57436 is that a sinusoid with frequency ω is mapped to a sinusoid with frequency $T\omega$ where $T>1$ is an integer defining the order of the transposition. An attractive feature of the harmonic transposition is that it stretches a source frequency range into a target frequency range by a factor equal to the order of transposition, i.e. by a factor equal to T. The harmonic transposition performs well for complex musical material. Furthermore, harmonic transposition exhibits low cross over frequencies, i.e. a large high frequency range above the cross over frequency can be generated from a relatively small low frequency range below the cross over frequency.

In contrast to harmonic transposition, a single sideband modulation (SSB) based HFR maps a sinusoid with frequency ω to a sinusoid with frequency $\omega+\Delta\omega$ where $\Delta\omega$ is a fixed frequency shift. It has been observed that, given a core signal with low bandwidth, a dissonant ringing artifact may result from the SSB transposition. It should also be noted that for a low cross-over frequency, i.e. a small source frequency range, harmonic transposition will require a smaller number of patches in order to fill a desired target frequency range than SSB based transposition. By way of example, if the high frequency range of $(\omega,4\omega]$ should be filled, then using an order of transposition $T=4$ harmonic transposition can fill this frequency range from a low frequency range of $(\frac{1}{4}\omega,\omega]$. On the other hand, a SSB based transposition using the same low frequency range must use a

frequency shift of $\Delta\omega = \frac{3}{4}\omega$ and it is necessary to repeat the process four times in order to fill the high frequency range $(\omega, 4\omega]$.

On the other hand, as already pointed out in WO 02/052545 A1, harmonic transposition has drawbacks for signals with a prominent periodic structure. Such signals are superimpositions of harmonically related sinusoids with frequencies $\Omega, 2\Omega, 3\Omega, \ldots$, where Ω is the fundamental frequency.
Upon harmonic transposition of order T, the output sinusoids have frequencies $T\Omega, 2T\Omega, 3T\Omega, \ldots$, which, in case of $T > 1$, is only a strict subset of the desired full harmonic series. In terms of resulting audio quality a "ghost" pitch corresponding to the transposed fundamental frequency $T\Omega$ will typically be perceived. Often the harmonic transposition results in a "metallic" sound character of the encoded and decoded audio signal. The situation may be alleviated to a certain degree by adding several orders of transposition $T = 2, 3, \ldots, T_{\max}$ to the HFR, but this method is computationally complex if most spectral gaps are to be avoided.

An alternative solution for avoiding the appearance of "ghost" pitches when using harmonic transposition has been presented in WO 02/052545 A1. The solution consists in using two types of transposition, i.e. a typical harmonic transposition and a special "pulse transposition". The described method teaches to switch to the dedicated "pulse transposition" for parts of the audio signal that are detected to be periodic with pulse-train like character. The problem with this approach is that the application of "pulse transposition" on complex music material often degrades the quality compared to harmonic transposition based on a high resolution filter bank. Hence, the detection mechanisms have to be tuned rather conservatively such that pulse transposition is not used for complex material. Inevitably, single pitch instruments and voices will sometimes be classified as complex signals, hereby invoking harmonic transposition and therefore missing harmonics. Moreover, if switching occurs in the middle of a single pitched signal, or a signal with a dominating pitch in a weaker complex background, the switching itself between the two transposition methods having very different spectrum filling properties will generate audible artifacts.

SUMMARY OF THE INVENTION

The present invention provides a method and system to complete the harmonic series resulting from harmonic transposition of a periodic signal. Frequency domain transposition comprises the step of mapping nonlinearly modified subband signals from an analysis filter bank into selected subbands of a synthesis filter bank. The nonlinear modification comprises a phase modification or phase rotation which in a complex filter bank domain can be obtained by a power law followed by a magnitude adjustment. Whereas prior art transposition modifies one analysis subband at a time separately, the present invention teaches to add a nonlinear combination of at least two different analysis subbands for each synthesis subband. The spacing between the analysis subbands to be combined may be related to the fundamental frequency of a dominant component of the signal to be transposed.

In the most general form, the mathematical description of the invention is that a set of frequency components $\omega_1, \omega_2, \ldots, \omega_K$ are used to create a new frequency component

$$\omega = T_1\omega_1 + T_2\omega_2 + \ldots + T_K\omega_K ,$$

where the coefficients $T_1, T_2 \ldots, T_K$ are integer transposition orders whose sum is the total transposition order $T = T_1 + T_2 + \ldots + T_K$. This effect is obtained by modifying the phases of K suitably chosen subband signals by the factors $T_1, T_2 \ldots, T_K$ and recombining the result into a signal with phase equal to the sum of the modified phases. It is important to note that all these phase operations are well defined and unambiguous since the individual transposition orders are integers, and that some of these integers could even be negative as long as the total transposition order satisfies $T \geq 1$.

The prior art methods correspond to the case $K = 1$, and the current invention teaches to use $K \geq 2$. The descriptive text treats mainly the case $K = 2$, $T \geq 2$ as it is sufficient to solve most specific problems at hand. But it should be noted that the cases $K > 2$ are considered to be equally disclosed and covered by the present document.

The invention uses information from a higher number of lower frequency band analytical channels, i.e. a higher number of analysis subband signals, to map the nonlinearly

modified subband signals from an analysis filter bank into selected sub-bands of a synthesis filter bank. The transposition is not just modifying one sub-band at a time separately but it adds a nonlinear combination of at least two different analysis sub-bands for each synthesis sub-band. As already mentioned, harmonic transposition of order T is designed to map a sinusoid of frequency ω to a sinusoid with frequency $T\omega$, with $T>1$. According to the invention, a so-called cross product enhancement with pitch parameter Ω and an index $0<r<T$ is designed to map a *pair* of sinusoids with frequencies $(\omega,\omega+\Omega)$ to a sinusoid with frequency $(T-r)\omega+r(\omega+\Omega)=T\omega+r\Omega$. It should be appreciated that for such cross product transpositions all partial frequencies of a periodic signal with a period of Ω will be generated by adding all cross products of pitch parameter Ω, with the index r ranging from 1 to $T-1$, to the harmonic transposition of order T.

According to an aspect of the invention, a system and a method for generating a high frequency component of a signal from a low frequency component of the signal is described. It should be noted that the features described in the following in the context of a system are equally applicable to the inventive method. The signal may e.g. be an audio and/or a speech signal. The system and method may be used for unified speech and audio signal coding. The signal comprises a low frequency component and a high frequency component, wherein the low frequency component comprises the frequencies below a certain cross-over frequency and the high frequency component comprises the frequencies above the cross-over frequency. In certain circumstances it may be required to estimate the high frequency component of the signal from its low frequency component. By way of example, certain audio encoding schemes only encode the low frequency component of an audio signal and aim at reconstructing the high frequency component of that signal solely from the decoded low frequency component, possibly by using certain information on the envelope of the original high frequency component. The system and method described here may be used in the context of such encoding and decoding systems.

The system for generating the high frequency component comprises an analysis filter bank which provides a plurality of analysis subband signals of the low frequency component of the signal. Such analysis filter banks may comprise a set of bandpass filters with constant bandwidth. Notably in the context of speech signals, it may also be

beneficial to use a set of bandpass filters with a logarithmic bandwidth distribution. It is an aim of the analysis filter bank to split up the low frequency component of the signal into its frequency constituents. These frequency constituents will be reflected in the plurality of analysis subband signals generated by the analysis filter bank. By way of example, a signal comprising a note played by musical instrument will be split up into analysis subband signals having a significant magnitude for subbands that correspond to the harmonic frequency of the played note, whereas other subbands will show analysis subband signals with low magnitude.

The system comprises further a non-linear processing unit to generate a synthesis subband signal with a particular synthesis frequency by modifying or rotating the phase of a first and a second of the plurality of analysis subband signals and by combining the phase-modified analysis subband signals. The first and the second analysis subband signals are different, in general. In other words, they correspond to different subbands. The non-linear processing unit may comprise a so-called cross-term processing unit within which the synthesis subband signal is generated. The synthesis subband signal comprises the synthesis frequency. In general, the synthesis subband signal comprises frequencies from a certain synthesis frequency range. The synthesis frequency is a frequency within this frequency range, e.g. a center frequency of the frequency range. The synthesis frequency and also the synthesis frequency range are typically above the cross-over frequency. In an analogous manner the analysis subband signals comprise frequencies from a certain analysis frequency range. These analysis frequency ranges are typically below the cross-over frequency.

The operation of phase modification may consist in transposing the frequencies of the analysis subband signals. Typically, the analysis filter bank yields complex analysis subband signals which may be represented as complex exponentials comprising a magnitude and a phase. The phase of the complex subband signal corresponds to the frequency of the subband signal. A transposition of such subband signals by a certain transposition order T' may be performed by taking the subband signal to the power of the transposition order T'. This results in the phase of the complex subband signal to be multiplied by the transposition order T'. By consequence, the transposed analysis subband signal exhibits a phase or a frequency which is T' times greater than the initial

phase or frequency. Such phase modification operation may also be referred to as phase rotation or phase multiplication.

The system comprises, in addition, a synthesis filter bank for generating the high frequency component of the signal from the synthesis subband signal. In other words, the aim of the synthesis filter bank is to merge possibly a plurality of synthesis subband signals from possibly a plurality of synthesis frequency ranges and to generate a high frequency component of the signal in the time domain. It should be noted that for signals comprising a fundamental frequency, e.g. a fundamental frequency Ω, it may be beneficial that the synthesis filter bank and/or the analysis filter bank exhibit a frequency spacing which is associated with the fundamental frequency of the signal. In particular, it may be beneficial to choose filter banks with a sufficiently low frequency spacing or a sufficiently high resolution in order to resolve the fundamental frequency Ω.

According to another aspect of the invention, the non-linear processing unit or the cross-term processing unit within the non-linear processing unit comprises a multiple-input-single-output unit of a first and second transposition order generating the synthesis subband signal from the first and the second analysis subband signal exhibiting a first and a second analysis frequency, respectively. In other words, the multiple-input-single-output unit performs the transposition of the first and second analysis subband signals and merges the two transposed analysis subband signals into a synthesis subband signal. The first analysis subband signal is phase-modified, or its phase is multiplied, by the first transposition order and the second analysis subband signal is phase-modified, or its phase is multiplied, by the second transposition order. In case of complex analysis subband signals such phase modification operation consists in multiplying the phase of the respective analysis subband signal by the respective transposition order. The two transposed analysis subband signals are combined in order to yield a combined synthesis subband signal with a synthesis frequency which corresponds to the first analysis frequency multiplied by the first transposition order plus the second analysis frequency multiplied by the second transposition order. This combination step may consist in the multiplication of the two transposed complex analysis subband signals. Such multiplication between two signals may consist in the multiplication of their samples.

The above mentioned features may also be expressed in terms of formulas. Let the first analysis frequency be ω and the second analysis frequency be $(\omega+\Omega)$. It should be noted that these variables may also represent the respective analysis frequency ranges of the two analysis subband signals. In other words, a frequency should be understood as representing all the frequencies comprised within a particular frequency range or frequency subband, i.e. the first and second analysis frequency should also be understood as a first and a second analysis frequency range or a first and a second analysis subband. Furthermore, the first transposition order may be (T-r) and the second transposition order may be r. It may be beneficial to restrict the transposition orders such that $T>1$ and $1 \leq r < T$. For such cases the multiple-input-single-output unit may yield synthesis subband signals with a synthesis frequency of $(T-r)\cdot\omega + r\cdot(\omega+\Omega)$.

According to a further aspect of the invention, the system comprises a plurality of multiple-input-single-output units and/or a plurality of non-linear processing units which generate a plurality of partial synthesis subband signals having the synthesis frequency. In other words, a plurality of partial synthesis subband signals covering the same synthesis frequency range may be generated. In such cases, a subband summing unit is provided for combining the plurality of partial synthesis subband signals. The combined partial synthesis subband signals then represent the synthesis subband signal. The combining operation may comprise the adding up of the plurality of partial synthesis subband signals. It may also comprise the determination of an average synthesis subband signal from the plurality of partial synthesis subband signals, wherein the synthesis subband signals may be weighted according to their relevance for the synthesis subband signal. The combining operation may also comprise the selecting of one or some of the plurality of subband signals which e.g. have a magnitude which exceeds a pre-defined threshold value. It should be noted that it may be beneficial that the synthesis subband signal is multiplied by a gain parameter. Notably in cases, where there is a plurality of partial synthesis subband signals, such gain parameters may contribute to the normalization of the synthesis subband signals.

According to a further aspect of the invention, the non-linear processing unit further comprises a direct processing unit for generating a further synthesis subband signal from a third of the plurality of analysis subband signals. Such direct processing unit may execute the direct transposition methods described e.g. in WO 98/57436. If the system

comprises an additional direct processing unit, then it may be necessary to provide a subband summing unit for combining corresponding synthesis subband signals. Such corresponding synthesis subband signals are typically subband signals covering the same synthesis frequency range and/or exhibiting the same synthesis frequency. The subband summing unit may perform the combination according to the aspects outlined above. It may also ignore certain synthesis subband signals, notably the once generated in the multiple-input-single-output units, if the minimum of the magnitude of the one or more analysis subband signals, e.g. from the cross-terms contributing to the synthesis subband signal, are smaller than a pre-defined fraction of the magnitude of the signal. The signal may be the low frequency component of the signal or a particular analysis subband signal. This signal may also be a particular synthesis subband signal. In other words, if the energy or magnitude of the analysis subband signals used for generating the synthesis subband signal is too small, then this synthesis subband signal may not be used for generating a high frequency component of the signal. The energy or magnitude may be determined for each sample or it may be determined for a set of samples, e.g. by determining a time average or a sliding window average across a plurality of adjacent samples, of the analysis subband signals.

The direct processing unit may comprise a single-input-single-output unit of a third transposition order T', generating the synthesis subband signal from the third analysis subband signal exhibiting a third analysis frequency, wherein the third analysis subband signal is phase-modified, or its phase is multiplied, by the third transposition order T' and wherein T' is greater than one. The synthesis frequency then corresponds to the third analysis frequency multiplied by the third transposition order. It should be noted that this third transposition order T' is preferably equal to the system transposition order T introduced below.

According to another aspect of the invention, the analysis filter bank has N analysis subbands at an essentially constant subband spacing of $\Delta\omega$. As mentioned above, this subband spacing $\Delta\omega$ may be associated with a fundamental frequency of the signal. An analysis subband is associated with an analysis subband index n, where $n \in \{1,...,N\}$. In other words, the analysis subbands of the analysis filter bank may be identified by a subband index n. In a similar manner, the analysis subband signals comprising

frequencies from the frequency range of the corresponding analysis subband may be identified with the subband index n.

On the synthesis side, the synthesis filter bank has a synthesis subband which is also associated with a synthesis subband index n. This synthesis subband index n also identifies the synthesis subband signal which comprises frequencies from the synthesis frequency range of the synthesis subband with subband index n. If the system has a system transposition order, also referred to as the total transposition order, T, then the synthesis subbands typically have an essentially constant subband spacing of $\Delta\omega \cdot T$, i.e. the subband spacing of the synthesis subbands is T times greater than the subband spacing of the analysis subbands. In such cases, the synthesis subband and the analysis subband with index n each comprise frequency ranges which relate to each other through the factor or the system transposition order T. By way of example, if the frequency range of the analysis subband with index n is $[(n-1)\cdot\omega, n\cdot\omega]$, then the frequency range of the synthesis subband with index n is $[T\cdot(n-1)\cdot\omega, T\cdot n\cdot\omega]$.

Given that the synthesis subband signal is associated with the synthesis subband with index n, another aspect of the invention is that this synthesis subband signal with index n is generated in a multiple-input-single-output unit from a first and a second analysis subband signal. The first analysis subband signal is associated with an analysis subband with index $n-p_1$ and the second analysis subband signal is associated with an analysis subband with index $n+p_2$.

In the following, several methods for selecting a pair of index shifts (p_1, p_2) are outlined. This may be performed by a so-called index selection unit. Typically, an optimal pair of index shifts is selected in order to generate a synthesis subband signal with a pre-defined synthesis frequency. In a first method, the index shifts p_1 and p_2 are selected from a limited list of pairs (p_1, p_2) stored in an index storing unit. From this limited list of index shift pairs, a pair (p_1, p_2) could be selected such that the minimum value of a set comprising the magnitude of the first analysis subband signal and the magnitude of the second analysis subband signal is maximized. In other words, for each possible pair of index shifts p_1 and p_2 the magnitude of the corresponding analysis subband signals could be determined. In case of complex analysis subband signals, the magnitude corresponds to the absolute value. The magnitude may be determined for each sample or it may be

determined for a set of samples, e.g. by determining a time average or a sliding window average across a plurality of adjacent samples, of the analysis subband signal. This yields a first and a second magnitude for the first and second analysis subband signal, respectively. The minimum of the first and the second magnitude is considered and the index shift pair (p_1, p_2) is selected for which this minimum magnitude value is highest.

In another method, the index shifts p_1 and p_2 are selected from a limited list of pairs (p_1, p_2), wherein the limited list is determined through the formulas $p_1 = r \cdot l$ and $p_2 = (T-r) \cdot l$. In these formulas l is a positive integer, taking on values e.g. from 1 to 10. This method is particularly useful in situations where the first transposition order used to transpose the first analysis subband ($n-p_1$) is (T-r) and where the second transposition order used to transpose the second analysis subband ($n+p_2$) is r. Assuming that the system transposition order T is fixed, the parameters l and r may be selected such that the minimum value of a set comprising the magnitude of the first analysis subband signal and the magnitude of the second analysis subband signal is maximized. In other words, the parameters l and r may be selected by a max-min optimization approach as outlined above.

In a further method, the selection of the first and second analysis subband signals may be based on characteristics of the underlying signal. Notably, if the signal comprises a fundamental frequency Ω, i.e. if the signal is periodic with pulse-train like character, it may be beneficial to select the index shifts p_1 and p_2 in consideration of such signal characteristic. The fundamental frequency Ω may be determined from the low frequency component of the signal or it may be determined from the original signal, comprising both, the low and the high frequency component. In the first case, the fundamental frequency Ω could be determined at a signal decoder using high frequency reconstruction, while in the second case the fundamental frequency Ω would typically be determined at a signal encoder and then signaled to the corresponding signal decoder. If an analysis filter bank with a subband spacing of $\Delta\omega$ is used and if the first transposition order used to transpose the first analysis subband ($n-p_1$) is (T-r) and if the second transposition order used to transpose the second analysis subband ($n+p_2$) is r then p_1 and p_2 may be selected such that their sum p_1+p_2 approximates the fraction $\Omega/\Delta\omega$ and their fraction p_1/p_2 approximates r/(T-r). In a particular case, p_1 and p_2 are selected such

that the fraction p_1/p_2 equals r/(T-r).

According to another aspect of the invention, the system for generating a high frequency component of a signal also comprises an analysis window which isolates a pre-defined time interval of the low frequency component around a pre-defined time instance k. The system may also comprise a synthesis window which isolates a pre-defined time interval of the high frequency component around a pre-defined time instance k. Such windows are particularly useful for signals with frequency constituents which are changing over time. They allow analyzing the momentary frequency composition of a signal. In combination with the filter banks a typical example for such time-dependent frequency analysis is the Short Time Fourier Transform (STFT). It should be noted that often the analysis window is a time-spread version of the synthesis window. For a system with a system order transposition T, the analysis window in the time domain may be a time spread version of the synthesis window in the time domain with a spreading factor T.

According to a further aspect of the invention, a system for decoding a signal is described. The system takes an encoded version of the low frequency component of a signal and comprises a transposition unit, according to the system described above, for generating the high frequency component of the signal from the low frequency component of the signal. Typically such decoding systems further comprise a core decoder for decoding the low frequency component of the signal. The decoding system may further comprise an upsampler for performing an upsampling of the low frequency component to yield an upsampled low frequency component. This may be required, if the low frequency component of the signal has been down-sampled at the encoder, exploiting the fact that the low frequency component only covers a reduced frequency range compared to the original signal. In addition, the decoding system may comprise an input unit for receiving the encoded signal, comprising the low frequency component, and an output unit for providing the decoded signal, comprising the low and the generated high frequency component.

The decoding system may further comprise an envelope adjuster to shape the high frequency component. While the high frequencies of a signal may be re-generated from the low frequency range of a signal using the high frequency reconstruction systems and methods described in the present document, it may be beneficial to extract information

from the original signal regarding the spectral envelope of its high frequency component. This envelope information may then be provided to the decoder, in order to generate a high frequency component which approximates well the spectral envelope of the high frequency component of the original signal. This operation is typically performed in the envelope adjuster at the decoding system. For receiving information related to the envelope of the high frequency component of the signal, the decoding system may comprise an envelope data reception unit. The regenerated high frequency component and the decoded and possibly upsampled low frequency component may then be summed up in a component summing unit to determine the decoded signal.

As outlined above, the system for generating the high frequency component may use information with regards to the analysis subband signals which are to be transposed and combined in order to generate a particular synthesis subband signal. For this purpose, the decoding system may further comprise a subband selection data reception unit for receiving information which allows the selection of the first and second analysis subband signals from which the synthesis subband signal is to be generated. This information may be related to certain characteristics of the encoded signal, e.g. the information may be associated with a fundamental frequency Ω of the signal. The information may also be directly related to the analysis subbands which are to be selected. By way of example, the information may comprise a list of possible pairs of first and second analysis subband signals or a list of pairs (p_1, p_2) of possible index shifts.

According to another aspect of the invention an encoded signal is described. This encoded signal comprises information related to a low frequency component of the decoded signal, wherein the low frequency component comprises a plurality of analysis subband signals. Furthermore, the encoded signal comprises information related to which two of the plurality of analysis subband signals are to be selected to generate a high frequency component of the decoded signal by transposing the selected two analysis subband signals. In other words, the encoded signal comprises a possibly encoded version of the low frequency component of a signal. In addition, it provides information, such as a fundamental frequency Ω of the signal or a list of possible index shift pairs (p_1,p_2), which will allow a decoder to regenerate the high frequency component of the signal based on the cross product enhanced harmonic transposition method outlined in

the present document.

According to a further aspect of the invention, a system for encoding a signal is described. This encoding system comprises a splitting unit for splitting the signal into a low frequency component and into a high frequency component and a core encoder for encoding the low frequency component. It also comprises a frequency determination unit for determining a fundamental frequency Ω of the signal and a parameter encoder for encoding the fundamental frequency Ω, wherein the fundamental frequency Ω is used in a decoder to regenerate the high frequency component of the signal. The system may also comprise an envelope determination unit for determining the spectral envelope of the high frequency component and an envelope encoder for encoding the spectral envelope. In other words, the encoding system removes the high frequency component of the original signal and encodes the low frequency component by a core encoder, e.g. an AAC or Dolby D encoder. Furthermore, the encoding system analyzes the high frequency component of the original signal and determines a set of information that is used at the decoder to regenerate the high frequency component of the decoded signal. The set of information may comprise a fundamental frequency Ω of the signal and/or the spectral envelope of the high frequency component.

The encoding system may also comprise an analysis filter bank providing a plurality of analysis subband signals of the low frequency component of the signal. Furthermore, it may comprise a subband pair determination unit for determining a first and a second subband signal for generating a high frequency component of the signal and an index encoder for encoding index numbers representing the determined first and the second subband signal. In other words, the encoding system may use the high frequency reconstruction method and/or system described in the present document in order to determine the analysis subbands from which high frequency subbands and ultimately the high frequency component of the signal may be generated. The information on these subbands, e.g. a limited list of index shift pairs (p_1,p_2), may then be encoded and provided to the decoder.

As highlighted above, the invention also encompasses methods for generating a high frequency component of a signal, as well as methods for decoding and encoding signals. The features outlined above in the context of systems are equally applicable to

corresponding methods. In the following selected aspects of the methods according to the invention are outlined. In a similar manner these aspects are also applicable to the systems outlined in the present document.

According to another aspect of the invention, a method for performing high frequency reconstruction of a high frequency component from a low frequency component of a signal is described. This method comprises the step of providing a first subband signal of the low frequency component from a first frequency band and a second subband signal of the low frequency component from a second frequency band. In other words, two subband signals are isolated from the low frequency component of the signal, the first subband signal encompasses a first frequency band and the second subband signal encompasses a second frequency band. The two frequency subbands are preferably different. In a further step, the first and the second subband signals are transposed by a first and a second transposition factor, respectively. The transposition of each subband signal may be performed according to known methods for transposing signals. In case of complex subband signals, the transposition may be performed by modifying the phase, or by multiplying the phase, by the respective transposition factor or transposition order. In a further step, the transposed first and second subband signals are combined to yield a high frequency component which comprises frequencies from a high frequency band.

The transposition may be performed such that the high frequency band corresponds to the sum of the first frequency band multiplied by the first transposition factor and the second frequency band multiplied by the second transposition factor. Furthermore, the transposing step may comprise the steps of multiplying the first frequency band of the first subband signal with the first transposition factor and of multiplying the second frequency band of the second subband signal with the second transposition factor. To simplify the explanation and without limiting its scope, the invention is illustrated for transposition of individual frequencies. It should be noted, however, that the transposition is performed not only for individual frequencies, but also for entire frequency bands, i.e. for a plurality of frequencies comprised within a frequency band. As a matter of fact, the transposition of frequencies and the transposition of frequency bands should be understood as being interchangeable in the present document. However, one has to be aware of different frequency resolutions of the analysis and

synthesis filterbanks.

In the above mentioned method, the providing step may comprise the filtering of the low frequency component by an analysis filter bank to generate a first and a second subband signal. On the other side, the combining step may comprise multiplying the first and the second transposed subband signals to yield a high subband signal and inputting the high subband signal into a synthesis filter bank to generate the high frequency component. Other signal transformations into and from a frequency representation are also possible and within the scope of the invention. Such signal transformations comprise Fourier Transforms (FFT, DCT), wavelet transforms, quadrature mirror filters (QMF), etc.. Furthermore, these transforms also comprise window functions for the purpose of isolating a reduced time interval of the "to be transformed" signal. Possible window functions comprise Gaussian windows, cosine windows, Hamming windows, Hann windows, rectangular windows, Barlett windows, Blackman windows, and others. In this document the term "filter bank" may comprise any such transforms possibly combined with any such window functions.

According to another aspect of the invention, a method for decoding an encoded signal is described. The encoded signal is derived from an original signal and represents only a portion of frequency subbands of the original signal below a cross-over frequency. The method comprises the steps of providing a first and a second frequency subband of the encoded signal. This may be done by using an analysis filter bank. Then the frequency subbands are transposed by a first transposition factor and a second transposition factor, respectively. This may be done by performing a phase modification, or a phase multiplication, of the signal in the first frequency subband with the first transposition factor and by performing a phase modification, or a phase multiplication, of the signal in the second frequency subband with the second transposition factor. Finally, a high frequency subband is generated from the first and second transposed frequency subbands, wherein the high frequency subband is above the cross-over frequency. This high frequency subband may correspond to the sum of the first frequency subband multiplied by the first transposition factor and the second frequency subband multiplied by the second transposition factor.

According to another aspect of the invention, a method for encoding a signal is described. This method comprises of the steps of filtering the signal to isolate a low frequency of the signal and of encoding the low frequency component of the signal. Furthermore, a plurality of analysis subband signals of the low frequency component of the signal is provided. This may be done using an analysis filter bank as described in the present document. Then a first and a second subband signal for generating a high frequency component of the signal are determined. This may be done using the high frequency reconstruction methods and systems outlined in the present document. Finally, information representing the determined first and the second subband signal is encoded. Such information may be characteristics of the original signal, e.g. the fundamental frequency Ω of the signal, or information related to the selected analysis subbands, e.g. the index shift pairs (p_1,p_2).

It should be noted that the above mentioned embodiments and aspects of the invention may be arbitrarily combined. In particular, it should be noted that the aspects outlined for a system are also applicable to the corresponding method embraced by the present invention. Furthermore, it should be noted that the disclosure of the invention also covers other claim combinations than the claim combinations which are explicitly given by the back references in the dependent claims, i.e., the claims and their technical features can be combined in any order and any formation.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described by way of illustrative examples, not limiting the scope of the invention. It will be described with reference to the accompanying drawings, in which:

Fig. 1 illustrates the operation of an HFR enhanced audio decoder;
Fig. 2 illustrates the operation of a harmonic transposer using several orders;
Fig. 3 illustrates the operation of a frequency domain (FD) harmonic transposer;
Fig. 4 illustrates the operation of the inventive use of cross term processing;
Fig. 5 illustrates prior art direct processing;
Fig. 6 illustrates prior art direct nonlinear processing of a single sub-band;

Fig. 7 illustrates the components of the inventive cross term processing;

Fig. 8 illustrates the operation of a cross term processing block;

Fig. 9 illustrates the inventive nonlinear processing contained in each of the MISO systems of Fig. 8;

Figs. 10 - 18 illustrate the effect of the invention for the harmonic transposition of exemplary periodic signals;

Fig. 19 illustrates the time-frequency resolution of a Short Time Fourier Transform (STFT);

Fig. 20 illustrates the exemplary time progression of a window function and its Fourier transform used on the synthesis side;

Fig. 21 illustrates the STFT of a sinusoidal input signal;

Fig. 22 illustrates the window function and its Fourier transform according to Fig. 20 used on the analysis side;

Figs. 23 and 24 illustrate the determination of appropriate analysis filter bank subbands for the cross-term enhancement of a synthesis filter band subband;

Figs. 25, 26, and 27 illustrate experimental results of the described direct-term and cross-term harmonic transposition method;

Figs. 28 and 29 illustrate embodiments of an encoder and a decoder, respectively, using the enhanced harmonic transposition schemes outlined in the present document; and

Fig. 30 illustrates an embodiment of a transposition unit shown in Figs. 28 and 29.

DESCRIPTION OF PREFERRED EMBODIMENTS

The below-described embodiments are merely illustrative for the principles of the present invention for the so-called CROSS PRODUCT ENHANCED HARMONIC TRANSPOSITION. 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.

Fig. 1 illustrates the operation of an HFR enhanced audio decoder. The core audio decoder 101 outputs a low bandwidth audio signal which is fed to an upsampler 104 which may be required in order to produce a final audio output contribution at the desired full sampling rate. Such upsampling is required for dual rate systems, where the band limited core audio codec is operating at half the external audio sampling rate, while

the HFR part is processed at the full sampling frequency. Consequently, for a single rate system, this upsampler 104 is omitted. The low bandwidth output of 101 is also sent to the transposer or the transposition unit 102 which outputs a transposed signal, i.e. a signal comprising the desired high frequency range. This transposed signal may be shaped in time and frequency by the envelope adjuster 103. The final audio output is the sum of low bandwidth core signal and the envelope adjusted transposed signal.

Fig. 2 illustrates the operation of a harmonic transposer 201, which corresponds to the transposer 102 of Fig. 1, comprising several transposers of different transposition order T. The signal to be transposed is passed to the bank of individual transposers 201-2, 201-3, ... , 201-T_{max} having orders of transposition $T = 2,3,\ldots,T_{max}$, respectively. Typically a transposition order $T_{max} = 3$ suffices for most audio coding applications. The contributions of the different transposers 201-2, 201-3, ... , 201-T_{max} are summed in 202 to yield the combined transposer output. In a first embodiment, this summing operation may comprise the adding up of the individual contributions. In another embodiment, the contributions are weighted with different weights, such that the effect of adding multiple contributions to certain frequencies is mitigated. For instance, the third order contributions may be added with a lower gain than the second order contributions. Finally, the summing unit 202 may add the contributions selectively depending on the output frequency. For instance, the second order transposition may be used for a first lower target frequency range, and the third order transposition may be used for a second higher target frequency range.

Fig. 3 illustrates the operation of a frequency domain (FD) harmonic transposer, such as one of the individual blocks of 201, i.e. one of the transposers 201-T of transposition order T. An analysis filter bank 301 outputs complex subbands that are submitted to nonlinear processing 302, which modifies the phase and/or amplitude of the subband signal according to the chosen transposition order T. The modified subbands are fed to a synthesis filterbank 303 which outputs the transposed time domain signal. In the case of multiple parallel transposers of different transposition orders such as shown in Fig. 2, some filter bank operations may be shared between different transposers 201-2, 201-3, ... , 201-T_{max}. The sharing of filter bank operations may be done for analysis or synthesis. In the case of shared synthesis 303, the summing 202 can be performed in the subband domain, i.e. before the synthesis 303.

Fig. 4 illustrates the operation of cross term processing 402 in addition to the direct processing 401. The cross term processing 402 and the direct processing 401 are performed in parallel within the nonlinear processing block 302 of the frequency domain harmonic transposer of Fig. 3. The transposed output signals are combined, e.g. added, in order to provide a joint transposed signal. This combination of transposed output signals may consist in the superposition of the transposed output signals. Optionally, the selective addition of cross terms may be implemented in the gain computation.

Fig. 5 illustrates in more detail the operation of the direct processing block 401 of Fig. 4 within the frequency domain harmonic transposer of Fig. 3. Single-input-single-output (SISO) units 401-1, ... , 401-n, ... , 401-N map each analysis subband from a source range into one synthesis subband in a target range. According to the Fig. 5, an analysis subband of index n is mapped by the SISO unit 401-n to a synthesis subband of the same index *n*. It should be noted that the frequency range of the subband with index *n* in the synthesis filter bank may vary depending on the exact version or type of harmonic transposition. In the version or type illustrated in Fig. 5, the frequency spacing of the analysis bank 301 is a factor T smaller than that of the synthesis bank 303. Hence, the index n in the synthesis bank 303 corresponds to a frequency, which is T times higher than the frequency of the subband with the same index n in the analysis bank 301. By way of example, an analysis subband $[(n-1)\omega, n\omega]$ is transposed into a synthesis subband $[(n-1)T\omega, nT\omega]$.

Fig. 6 illustrates the direct nonlinear processing of a single subband contained in each of the SISO units of 401-n. The nonlinearity of block 601 performs a multiplication of the phase of the complex subband signal by a factor equal to the transposition order T. The optional gain unit 602 modifies the magnitude of the phase modified subband signal. In mathematical terms, the output y of the SISO unit 401-n can be written as a function of the input x to the SISO system 401-n and the gain parameter g as follows:

$$y = g \cdot v^T \text{, where } v = x/|x|^{1-1/T} . \quad (1)$$

This may also be written as:

$$y = g \cdot |x| \cdot \left(\frac{x}{|x|} \right)^T .$$

In words, the phase of the complex subband signal x is multiplied by the transposition order T and the amplitude of the complex subband signal x is modified by the gain parameter *g*.

Fig. 7 illustrates the components of the cross term processing 402 for an harmonic transposition of order T. There are $T-1$ cross term processing blocks in parallel, 701-1, ..., 701-r, ... 701-(T-1), whose outputs are summed in the summing unit 702 to produce a combined output. As already pointed out in the introductory section, it is a target to map a *pair* of sinusoids with frequencies $(\omega, \omega+\Omega)$ to a sinusoid with frequency $(T-r)\omega + r(\omega+\Omega) = T\omega + r\Omega$, wherein the variable r varies from 1 to $T-1$. In other words, two subbands from the analysis filter bank 301 are to be mapped to one subband of the high frequency range. For a particular value of r and a given transposition order T, this mapping step is performed in the cross term processing block 701-r.

Fig. 8 illustrates the operation of a cross term processing block 701-r for a fixed value $r = 1, 2, \ldots, T-1$. Each output subband 803 is obtained in a multiple-input-single-output (MISO) unit 800-n from two input subbands 801 and 802. For an output subband 803 of index n, the two inputs of the MISO unit 800-n are subbands $n - p_1$, 801, and $n + p_2$, 802, where p_1 and p_2 are positive integer index shifts, which depend on the transposition order T, the variable r, and the cross product enhancement pitch parameter Ω. The analysis and synthesis subband numbering convention is kept in line with that of Fig 5, that is, the spacing in frequency of the analysis bank 301 is a factor T smaller than that of the synthesis bank 303 and consequently the above comments given on variations of the factor T remain relevant.

In relation to the usage of cross term processing, the following remarks should be considered. The pitch parameter Ω does not have to be known with high precision, and certainly not with better frequency resolution than the frequency resolution obtained by the analysis filter bank 301. In fact, in some embodiments of the present invention, the

underlying cross product enhancement pitch parameter Ω is not entered in the decoder at all. Instead, the chosen pair of integer index shifts (p_1, p_2) is selected from a list of possible candidates by following an optimization criterion such as the maximization of the cross product output magnitude, i.e. the maximization of the energy of the cross product output. By way of example, for given values of T and r, a list of candidates given by the formula $(p_1, p_2) = (rl, (T-r)l), l \in L$, where L is a list of positive integers, could be used. This is shown in further detail below in the context of formula (11). All positive integers are in principle OK as candidates. In some cases pitch information may help to identify which *l* to choose as appropriate index shifts.

Furthermore, even though the example cross product processing illustrated in Fig. 8 suggests that the applied index shifts (p_1, p_2) are the same for a certain range of output subbands, e.g. synthesis subbands (n-1), n and (n+1) are composed from analysis subbands having a fixed distance $p_1 + p_2$, this need not be the case. As a matter of fact, the index shifts (p_1, p_2) may differ for each and every output subband. This means that for each subband *n* a different value Ω of the cross product enhancement pitch parameter may be selected.

Fig. 9 illustrates the nonlinear processing contained in each of the MISO units 800-n. The product operation 901 creates a subband signal with a phase equal to a weighted sum of the phases of the two complex input subband signals and a magnitude equal to a generalized mean value of the magnitudes of the two input subband samples. The optional gain unit 902 modifies the magnitude of the phase modified subband samples. In mathematical terms, the output y can be written as a function of the inputs u_1 801 and u_2 802 to the MISO unit 800-n and the gain parameter g as follows,

$$y = g \cdot v_1^{T-r} v_2^r \text{ , where } v_m = u_m / |u_m|^{1-1/T} \text{ , for } m = 1,2. \qquad (2)$$

This may also be written as:

$$y = \mu(|u_1|, |u_2|) \cdot \left(\frac{u_1}{|u_1|}\right)^{T-r} \left(\frac{u_2}{|u_2|}\right)^{r},$$

where $\mu(|u_1|,|u_2|)$ is a magnitude generation function. In words, the phase of the complex subband signal u_1 is multiplied by the transposition order $T-r$ and the phase of the complex subband signal u_2 is multiplied by the transposition order r. The sum of those two phases is used as the phase of the output y whose magnitude is obtained by the magnitude generation function. Comparing with the formula (2) the magnitude generation function is expressed as the geometric mean of magnitudes modified by the gain parameter g, that is $\mu(|u_1|,|u_2|) = g \cdot |u_1|^{1-r/T} |u_2|^{r/T}$. By allowing the gain parameter to depend on the inputs this of course covers all possibilities.

It should be noted that the formula (2) results from the underlying target that a pair of sinusoids with frequencies $(\omega, \omega+\Omega)$ are to be mapped to a sinusoid with frequency $T\omega + r\Omega$, which can also be written as $(T-r)\omega + r(\omega+\Omega)$.

In the following text, a mathematical description of the present invention will be outlined. For simplicity, continuous time signals are considered. The synthesis filter bank 303 is assumed to achieve perfect reconstruction from a corresponding complex modulated analysis filter bank 301 with a real valued symmetric window function or prototype filter $w(t)$. The synthesis filter bank will often, but not always, use the same window in the synthesis process. The modulation is assumed to be of an evenly stacked type, the stride is normalized to one and the angular frequency spacing of the synthesis subbands is normalized to π. Hence, a target signal $s(t)$ will be achieved at the output of the synthesis filter bank if the input subband signals to the synthesis filter bank are given by synthesis subband signals $y_n(k)$,

$$y_n(k) = \int_{-\infty}^{\infty} s(t) w(t-k) \exp\left[-in\pi(t-k)\right] dt \,. \qquad (3)$$

Note that formula (3) is a normalized continuous time mathematical model of the usual operations in a complex modulated subband analysis filter bank, such as a windowed Discrete Fourier Transform (DFT), also denoted as a Short Time Fourier Transform (STFT). With a slight modification in the argument of the complex exponential of formula (3), one obtains continuous time models for complex modulated (pseudo) Quadrature Mirror Filterbank (QMF) and complexified Modified Discrete Cosine Transform (CMDCT), also

denoted as a windowed oddly stacked windowed DFT. The subband index n runs through all nonnegative integers for the continuous time case. For the discrete time counterparts, the time variable t is sampled at step $1/N$, and the subband index n is limited by N, where N is the number of subbands in the filter bank, which is equal to the discrete time stride of the filter bank. In the discrete time case, a normalization factor related to N is also required in the transform operation if it is not incorporated in the scaling of the window.

For a real valued signal, there are as many complex subband samples out as there are real valued samples in for the chosen filter bank model. Therefore, there is a total oversampling (or redundancy) by a factor two. Filter banks with a higher degree of oversampling can also be employed, but the oversampling is kept small in the present description of embodiments for the clarity of exposition.

The main steps involved in the modulated filter bank analysis corresponding to formula (3) are that the signal is multiplied by a window centered around time $t = k$, and the resulting windowed signal is correlated with each of the complex sinusoids $\exp\left[-in\pi(t-k)\right]$. In discrete time implementations this correlation is efficiently implemented via a Fast Fourier Transform. The corresponding algorithmic steps for the synthesis filter bank are well known for those skilled in the art, and consist of synthesis modulation, synthesis windowing, and overlap add operations.

Fig. 19 illustrates the position in time and frequency corresponding to the information carried by the subband sample $y_n(k)$ for a selection of values of the time index k and the subband index n. As an example, the subband sample $y_5(4)$ is represented by the dark rectangle 1901.

For a sinusoid, $s(t) = A\cos(\omega t + \theta) = \mathrm{Re}\left\{C\exp(i\omega t)\right\}$, the subband signals of (3) are for sufficiently large n with good approximation given by

$$y_n(k) = Ce^{ik\omega}\int_{-\infty}^{\infty} w(t)\exp\left[-i(n\pi-\omega)t\right]dt = Ce^{ik\omega}\hat{w}(n\pi-\omega), \qquad (4)$$

where the hat denotes the Fourier transform, i.e. $\hat{w}$ is the Fourier transform of the window function w.

Strictly speaking, formula (4) is only true if one adds a term with -ω instead of ω. This term is neglected based on the assumption that the frequency response of the window decays sufficiently fast, and that the sum of ω and n is not close to zero.

Fig. 20 depicts the typical appearance of a window w, 2001, and its Fourier transform $\hat{w}$,2002.

Fig. 21 illustrates the analysis of a single sinusoid corresponding to formula (4). The subbands that are mainly affected by the sinusoid at frequency ω are those with index n such that $n\pi-\omega$ is small. For the example of Fig. 21, the frequency is $\omega=6.25\pi$ as indicated by the horizontal dashed line 2101. In that case, the three subbands for $n=5,6,7$, represented by reference signs 2102, 2103, 2104, respectively, contain significant nonzero subband signals. The shading of those three subbands reflects the relative amplitude of the complex sinusoids inside each subband obtained from formula (4). A darker shade means higher amplitude. In the concrete example, this means that the amplitude of subband 5, i.e. 2102, is lower compared to the amplitude of subband 7, i.e. 2104, which again is lower than the amplitude of subband 6, i.e. 2103. It is important to note that several nonzero subbands may in general be necessary to be able to synthesize a high quality sinusoid at the output of the synthesis filter bank, especially in cases where the window has an appearance like the window 2001 of Fig 20, with relatively short time duration and significant side lobes in frequency.

The synthesis subband signals $y_n(k)$ can also be determined as a result of the analysis filter bank 301 and the non-linear processing, i.e. harmonic transposer 302 illustrated in Fig. 3. On the analysis filter bank side, the analysis subband signals $x_n(k)$ may be represented as a function of the source signal $z(t)$. For a transposition of order T, a complex modulated analysis filter bank with window $w_T(t)=w(t/T)/T$, a stride one, and a modulation frequency step, which is T times finer than the frequency step of the synthesis bank, is applied on the source signal $z(t)$. Fig. 22 illustrates the appearance of

the scaled window w_T 2201 and its Fourier transform $\hat{w}_T$ 2202. Compared to Fig. 20, the time window 2201 is stretched out and the frequency window 2202 is compressed.

The analysis by the modified filter bank gives rise to the analysis subband signals $x_n(k)$:

$$x_n(k) = \int_{-\infty}^{\infty} z(t) w_T(t-k) \exp\left[-i\frac{n\pi}{T}(t-k)\right] dt \qquad (5)$$

For a sinusoid, $z(t) = B\cos(\xi t + \varphi) = \mathrm{Re}\{D\exp(i\xi t)\}$, one finds that the subband signals of (5) for sufficiently large n with good approximation are given by

$$x_n(k) = D\exp(ik\xi)\hat{w}(n\pi - T\xi). \qquad (6)$$

Hence, submitting these subband signals to the harmonic transposer 302 and applying the direct transposition rule (1) to (6) yields

$$\tilde{y}_n(k) = gD\left(\frac{D}{|D|}\right)^{T-1}\left(\frac{\hat{w}(n\pi - T\xi)}{|\hat{w}(n\pi - T\xi)|}\right)^{T-1} \cdot \exp(ikT\xi)\hat{w}(n\pi - T\xi). \qquad (7)$$

The synthesis subband signals $y_n(k)$ given by formula (4) and the nonlinear subband signals obtained through harmonic transposition $\tilde{y}_n(k)$ given by formal (7) ideally should match.

For odd transposition orders T, the factor containing the influence of the window in (7) is equal to one, since the Fourier transform of the window is real valued by assumption, and $T-1$ is an even number. Therefore, formula (7) can be matched exactly to formula (4) with $\omega = T\xi$, for all subbands, such that the output of the synthesis filter bank with input subband signals according to formula (7) is a sinusoid with a frequency $\omega = T\xi$, amplitude $A = gB$, and phase $\theta = T\varphi$, wherein B and φ are determined from the formula: $D = B\exp(i\varphi)$, which upon insertion yields $gD\left(\frac{D}{|D|}\right)^{T-1} = gB\exp(iT\varphi)$. Hence, a harmonic transposition of order T of the sinusoidal source signal $z(t)$ is obtained.

For even T, the match is more approximate, but it still holds on the positive valued part of the window frequency response $\hat{w}$, which for a symmetric real valued window includes the most important main lobe. This means that also for even values of T a harmonic transposition of the sinusoidal source signal $z(t)$ is obtained. In the particular case of a Gaussian window, $\hat{w}$ is always positive and consequently, there is no difference in performance for even and odd orders of transposition.

Similarly to formula (6), the analysis of a sinusoid with frequency $\xi+\Omega$, i.e. the sinusoidal source signal $z(t) = B'\cos((\zeta+\Omega)t+\varphi') = \mathrm{Re}\{E\exp(i(\zeta+\Omega)t)\}$, is

$$x'_n(k) = E\exp\big(ik(\xi+\Omega)\big)\hat{w}(n\pi - T(\xi+\Omega)). \tag{8}$$

Therefore, feeding the two subband signals $u_1 = x_{n-p_1}(k)$, which corresponds to the signal 801 in Fig. 8, and $u_2 = x'_{n+p_2}(k)$, which corresponds to the signal 802 in Fig. 8, into the cross product processing 800-n illustrated in Fig. 8 and applying the cross product formula (2) yields the output subband signal 803

$$\tilde{y}_n(k) = g\exp\Big[ik\big(T\xi + r\Omega\big)\Big]M(n,\xi), \tag{9}$$

where

$$M(n,\xi) = \frac{D^{T-r}E^r}{\left|D^{T-r}E^r\right|^{1-1/T}} \frac{\hat{w}\big((n-p_1)\pi - T\xi\big)^{T-r}\,\hat{w}\big((n+p_2)\pi - T(\xi+\Omega)\big)^r}{\left|\hat{w}\big((n-p_1)\pi - T\xi\big)^{T-r}\,\hat{w}\big((n+p_2)\pi - T(\xi+\Omega)\big)^r\right|^{1-1/T}}. \tag{10}$$

From formula (9) it can be seen that the phase evolution of the output subband signal 803 of the MISO system 800-n follows the phase evolution of an analysis of a sinusoid of frequency $T\xi + r\Omega$. This holds independently of the choice of the index shifts p_1 and p_2. In fact, if the subband signal (9) is fed into a subband channel n corresponding to the frequency $T\xi + r\Omega$, that is if $n\pi \approx T\xi + r\Omega$, then the output will be a contribution to the generation of a sinusoid at frequency $T\xi + r\Omega$. However, it is advantageous to make sure

that each contribution is significant, and that the contributions add up in a beneficial fashion. These aspects will be discussed below.

Given a cross product enhancement pitch parameter Ω, suitable choices for index shifts p_1 and p_2 can be derived in order for the complex magnitude $M(n,\xi)$ of (10) to approximate $\hat{w}\left(n\pi-\left(T\xi+r\Omega\right)\right)$ for a range of subbands n, in which case the final output will approximate a sinusoid at the frequency $T\xi+r\Omega$. A first consideration on main lobes imposes all three values of $\left(n-p_1\right)\pi-T\xi$, $\left(n+p_2\right)\pi-T\left(\xi+\Omega\right)$, $n\pi-\left(T\xi+r\Omega\right)$ to be small simultaneously, which leads to the approximate equalities

$$p_1 \approx r\frac{\Omega}{\pi} \text{ and } p_2 \approx (T-r)\frac{\Omega}{\pi}. \tag{11}$$

This means that when knowing the cross product enhancement pitch parameter Ω, the index shifts may be approximated by fomula (11), thereby allowing a simple selection of the analysis subbands. A more thorough analysis of the effects of the choice of the index shifts p_1 and p_2 according to formula (11) on the magnitude of the parameter $M(n,\zeta)$ according to formula (10) can be performed for important special cases of window functions $w(t)$ such as the Gaussian window and a sine window. One finds that the desired approximation to $\hat{w}\left(n\pi-\left(T\xi+r\Omega\right)\right)$ is very good for several subbands with $n\pi \approx T\xi+r\Omega$.

It should be noted that the relation (11) is calibrated to the exemplary situation where the analysis filter bank 301 has an angular frequency subband spacing of π/T. In the general case, the resulting interpretation of (11) is that the cross term source span p_1+p_2 is an integer approximating the underlying fundamental frequency Ω, measured in units of the analysis filter bank subband spacing, and that the pair (p_1,p_2) is chosen as a multiple of $(r,T-r)$.

For the determination of the index shift pair (p_1,p_2) in the decoder the following modes may be used:

1. A value of Ω may be derived in the encoding process and explicitly transmitted to the decoder in a sufficient precision to derive the integer values of p_1 and p_2 by means of a suitable rounding procedure, which may follow the principles that
   - $p_1 + p_2$ approximates $\Omega/\Delta\omega$, where $\Delta\omega$ is the angular frequency spacing of the analyis filter bank; and
   - p_1 / p_2 is chosen to approximate $r/(T-r)$.
2. For each target subband sample, the index shift pair (p_1, p_2) may be derived in the decoder from a pre-determined list of candidate values such as $(p_1, p_2) = (rl, (T-r)l), l \in L$, $r \in \{1,2,...,T-1\}$, where L is a list of positive integers. The selection may be based on an optimization of cross term output magnitude, e.g. a maximization of the energy of the cross term output.
3. For each target subband sample, the index shift pair (p_1, p_2) may be derived from a reduced list of candidate values by an optimization of cross term output magnitude, where the reduced list of candidate values is derived in the encoding process and transmitted to the decoder.

It should be noted that phase modification of the subband signals u_1 and u_2 is performed with a weighting $(T-r)$ and r, respectively, but the subband index distance p_1 and p_2 are chosen proportional to r and $(T-r)$, respectively. Thus the closest subband to the synthesis subband n receives the strongest phase modification.

An advantageous method for the optimization procedure for the modes 2 and 3 outlined above may be to consider the Max-Min optimization:

$$\max\left\{\min\left\{\left|x_{n-p_1}(k)\right|, \left|x_{n+p_2}(k)\right|\right\} : (p_1, p_2) = (rl, (T-r)l), l \in L, r \in \{1,2,...,T-1\}\right\}, \quad (12)$$

and to use the winning pair together with its corresponding value of r to construct the cross product contribution for a given target subband index n. In the decoder search oriented modes 2 and partially also 3, the addition of cross terms for different values r is preferably done independently, since there may be a risk of adding content to the same subband several times. If, on the other hand, the fundamental frequency Ω is used for selecting the subbands as in mode 1 or if only a narrow range of subband index distances

are permitted as may be the case in mode 2, this particular issue of adding content to the same subband several times may be avoided.

Furthermore, it should also be noted that for the embodiments of the cross term processing schemes outlined above an additional decoder modification of the cross product gain g may be beneficial. For instance, it is referred to the input subband signals u_1, u_2 to the cross products MISO unit given by formula (2) and the input subband signal x to the transposition SISO unit given by formula (1). If all three signals are to be fed to the same output synthesis subband as shown in Fig. 4, where the direct processing 401 and the cross product processing 402 provide components for the same output synthesis subband, it may be desirable to set the cross product gain g to zero, i.e. the gain unit 902 of Fig. 9, if

$$\min(|u_1|,|u_2|) < q|x|, \tag{13}$$

for a pre-defined threshold $q > 1$. In other words, the cross product addition is only performed if the direct term input subband magnitude $|x|$ is small compared to both of the cross product input terms. In this context, *x* is the analysis subband sample for the direct term processing which leads to an output at the same synthesis subband as the cross product under consideration. This may be a precaution in order to not enhance further a harmonic component that has already been furnished by the direct transposition.

In the following, the harmonic transposition method outlined in the present document will be described for exemplary spectral configurations to illustrate the enhancements over the prior art. Fig. 10 illustrates the effect of direct harmonic transposition of order $T = 2$. The top diagram 1001 depicts the partial frequency components of the original signal by vertical arrows positioned at multiples of the fundamental frequency Ω. It illustrates the source signal, e.g. at the encoder side. The diagram 1001 is segmented into a left sided source frequency range with the partial frequencies $\Omega, 2\Omega, 3\Omega, 4\Omega, 5\Omega$ and a right sided target frequency range with partial frequencies $6\Omega, 7\Omega, 8\Omega$. The source frequency range will typically be encoded and transmitted to the decoder. On the other hand, the right sided target frequency range, which comprises the partials $6\Omega, 7\Omega, 8\Omega$ above the cross

over frequency 1005 of the HFR method, will typically not be transmitted to the decoder. It is an object of the harmonic transposition method to reconstruct the target frequency range above the cross-over frequency 1005 of the source signal from the source frequency range. Consequently, the target frequency range, and notably the partials $6\Omega,7\Omega,8\Omega$ in diagram 1001 are not available as input to the transposer.

As outlined above, it is the aim of the harmonic transposition method to regenerate the signal components $6\Omega,7\Omega,8\Omega$ of the source signal from frequency components available in the source frequency range. The bottom diagram 1002 shows the output of the transposer in the right sided target frequency range. Such transposer may e.g. be placed at the decoder side. The partials at frequencies 6Ω and 8Ω are regenerated from the partials at frequencies 3Ω and 4Ω by harmonic transposition using an order of transposition $T = 2$. As a result of a spectral stretching effect of the harmonic transposition, depicted here by the dotted arrows 1003 and 1004, the target partial at 7Ω is missing. This target partial at 7Ω can not be generated using the underlying prior art harmonic transposition method.

Figure 11 illustrates the effect of the invention for harmonic transposition of a periodic signal in the case where a second order harmonic transposer is enhanced by a single cross term, i.e. $T = 2$ and $r = 1$. As outlined in the context of Fig. 10, a transposer is used to generate the partials $6\Omega,7\Omega,8\Omega$ in the target frequency range above the cross-over frequency 1105 in the lower diagram 1102 from the partials $\Omega,2\Omega,3\Omega,4\Omega,5\Omega$ in the source frequency range below the cross-over frequency 1105 of diagram 1101. In addition to the prior art transposer output of Figure 10, the partial frequency component at 7Ω is regenerated from a combination of the source partials at 3Ω and 4Ω. The effect of the cross product addition is depicted by dashed arrows 1103 and 1104. In terms of formulas, one has $\omega = 3\Omega$ and therefore $(T-r)\omega + r(\omega+\Omega) = T\omega + r\Omega = 6\Omega + \Omega = 7\Omega$. As can be seen from this example, all the target partials may be regenerated using the inventive HFR method outlined in the present document.

Fig. 12 illustrates a possible implementation of a prior art second order harmonic transposer in a modulated filter bank for the spectral configuration of Fig. 10. The stylized frequency responses of the analysis filter bank subbands are shown by dotted lines, e.g. reference sign 1206, in the top diagram 1201. The subbands are enumerated by the

subband index, of which the indexes 5, 10 and 15 are shown in Fig. 12. For the given example, the fundamental frequency Ω is equal to 3.5 times the analysis subband frequency spacing. This is illustrated by the fact that the partial Ω in diagram 1201 is positioned between the two subbands with subband index 3 and 4. The partial 2Ω is positioned in the center of the subband with subband index 7 and so forth.

The bottom diagram 1202 shows the regenerated partials 6Ω and 8Ω superimposed with the stylized frequency responses, e.g. reference sign 1207, of selected synthesis filter bank subbands. As described earlier, these subbands have a $T = 2$ times coarser frequency spacing. Correspondingly, also the frequency responses are scaled by the factor $T = 2$. As outlined above, the prior art direct term processing method modifies the phase of each analysis subband, i.e. of each subband below the cross-over frequency 1205 in diagram 1201, by a factor $T = 2$ and maps the result into the synthesis subband with the same index, i.e. a subband above the cross-over frequency 1205 in diagram 1202. This is symbolized in Fig. 12 by diagonal dotted arrows, e.g. arrow 1208 for the analysis subband 1206 and the synthesis subband 1207. The result of this direct term processing for subbands with subband indexes 9 to 16 from the analysis subband 1201 is the regeneration of the two target partials at frequencies 6Ω and 8Ω in the synthesis subband 1202 from the source partials at frequencies 3Ω and 4Ω. As can be seen from Fig. 12, the main contribution to the target partial 6Ω comes from the subbands with the subband indexes 10 and 11, i.e. reference signs 1209 and 1210, and the main contribution to the target partial 8Ω comes from the subband with subband index 14, i.e. reference sign 1211.

Fig. 13 illustrates a possible implementation of an additional cross term processing step in the modulated filter bank of Fig. 12. The cross-term processing step corresponds to the one described for periodic signals with the fundamental frequency Ω in relation to Fig. 11. The upper diagram 1301 illustrates the analysis subbands, of which the source frequency range is to be transposed into the target frequency range of the synthesis subbands in the lower diagram 1302. The particular case of the generation of the synthesis subbands 1315 and 1316, which are surrounding the partial 7Ω, from the analysis subbands is considered. For an order of transposition $T = 2$, a possible value $r = 1$ may be selected. Choosing the list of candidate values (p_1, p_2) as a multiple of

$(r, T-r) = (1,1)$ such that $p_1 + p_2$ approximates $\frac{\Omega}{\Delta\omega} = \frac{\Omega}{(\Omega/3.5)} = 3.5$, i.e. the fundamental frequency Ω in units of the analysis subband frequency spacing, leads to the choice $p_1 = p_2 = 2$. As outlined in the context of Fig. 8, a synthesis subband with the subband index *n* may be generated from the cross-term product of the analysis subbands with the subband index $(n - p_1)$ and $(n + p_2)$. Consequently, for the synthesis subband with subband index 12, i.e. reference sign 1315, a cross product is formed from the analysis subbands with subband index $(n - p_1) = 12 - 2 = 10$, i.e. reference sign 1311, and $(n + p_2) = 12 + 2 = 14$, i.e. reference sign 1313. For the synthesis subband with subband index 13, a cross product is formed from analysis subbands with and index $(n - p_1) = 13 - 2 = 11$, i.e. reference sign 1312, and $(n + p_2) = 13 + 2 = 15$, i.e. reference sign 1314. This process of cross-product generation is symbolized by the diagonal dashed/dotted arrow pairs, i.e. reference sign pairs 1308, 1309 and 1306, 1307, respectively.

As can be seen from Fig. 13, the partial 7Ω is placed primarily within the subband 1315 with index 12 and only secondarily in the subband 1316 with index 13. Consequently, for more realistic filter responses, there will be more direct and/or cross terms around synthesis subband 1315 with index 12 which add beneficially to the synthesis of a high quality sinusoid at frequency $(T-r)\omega + r(\omega + \Omega) = T\omega + r\Omega = 6\Omega + \Omega = 7\Omega$ than terms around synthesis subband 1316 with index 13. Furthermore, as highlighted in the context of formula (13), a blind addition of all cross terms with $p_1 = p_2 = 2$ could lead to unwanted signal components for less periodic and academic input signals. Consequently, this phenomenon of unwanted signal components may require the application of an adaptive cross product cancellation rule such as the rule given by formula (13).

Fig. 14 illustrates the effect of prior art harmonic transposition of order $T = 3$. The top diagram 1401 depicts the partial frequency components of the original signal by vertical arrows positioned at multiples of the fundamental frequency Ω. The partials $6\Omega, 7\Omega, 8\Omega, 9\Omega$ are in the target range above the cross over frequency 1405 of the HFR method and therefore not available as input to the transposer. The aim of the harmonic transposition is to regenerate those signal components from the signal in the source range. The bottom diagram 1402 shows the output of the transposer in the target

frequency range. The partials at frequencies 6Ω, i.e. reference sign 1407, and 9Ω, i.e. reference sign 1410, have been regenerated from the partials at frequencies 2Ω, i.e. reference sign 1406, and 3Ω, i.e. reference sign 1409. As a result of a spectral stretching effect of the harmonic transposition, depicted here by the dotted arrows 1408 and 1411, respectively, the target partials at 7Ω and 8Ω are missing.

Fig. 15 illustrates the effect of the invention for the harmonic transposition of a periodic signal in the case where a third order harmonic transposer is enhanced by the addition of two different cross terms, i.e. $T = 3$ and $r = 1,2$. In addition to the prior art transposer output of Fig. 14, the partial frequency component 1508 at 7Ω is regenerated by the cross term for $r = 1$ from a combination of the source partials 1506 at 2Ω and 1507 at 3Ω. The effect of the cross product addition is depicted by the dashed arrows 1510 and 1511. In terms of formulas, one has with $\omega = 2\Omega$, $(T-r)\omega + r(\omega + \Omega) = T\omega + r\Omega = 6\Omega + \Omega = 7\Omega$. Likewise, the partial frequency component 1509 at 8Ω is regenerated by the cross term for $r = 2$. This partial frequency component 1509 in the target range of the lower diagram 1502 is generated from the partial frequency components 1506 at 2Ω and 1507 at 3Ω in the source frequency range of the upper diagram 1501. The generation of the cross term product is depicted by the arrows 1512 and 1513. In terms of formulas, one has $(T-r)\omega + r(\omega + \Omega) = T\omega + r\Omega = 6\Omega + 2\Omega = 8\Omega$. As can be seen, all the target partials may be regenerated using the inventive HFR method described in the present document.

Fig. 16 illustrates a possible implementation of a prior art third order harmonic transposer in a modulated filter bank for the spectral situation of Fig. 14. The stylized frequency responses of the analysis filter bank subbands are shown by dotted lines in the top diagram 1601. The subbands are enumerated by the subband indexes 1 through 17 of which the subbands 1606, with index 7, 1607, with index 10 and 1608, with index 11, are referenced in an exemplary manner. For the given example, the fundamental frequency Ω is equal to 3.5 times the analysis subband frequency spacing $\Delta\omega$. The bottom diagram 1602 shows the regenerated partial frequency superimposed with the stylized frequency responses of selected synthesis filter bank subbands. By way of example, the subbands 1609, with subband index 7, 1610, with subband index 10 and 1611, with subband index 11 are referenced. As described above, these subbands have

a $T = 3$ times coarser frequency spacing $\Delta\omega$. Correspondingly, also the frequency responses are scaled accordingly.

The prior art direct term processing modifies the phase of the subband signals by a factor $T = 3$ for each analysis subband and maps the result into the synthesis subband with the same index, as symbolized by the diagonal dotted arrows. The result of this direct term processing for subbands 6 to 11 is the regeneration of the two target partial frequencies 6Ω and 9Ω from the source partials at frequencies 2Ω and 3Ω. As can be seen from Fig. 16, the main contribution to the target partial 6Ω comes from subband with index 7, i.e. reference sign 1606, and the main contributions to the target partial 9Ω comes from subbands with index 10 and 11, i.e. reference signs 1607 and 1608, respectively.

Fig. 17 illustrates a possible implementation of an additional cross term processing step for $r = 1$ in the modulated filter bank of Fig. 16 which leads to the regeneration of the partial at 7Ω. As was outlined in the context of Fig. 8 the index shifts (p_1, p_2) may be selected as a multiple of $(r, T-r) = (1,2)$, such that $p_1 + p_2$ approximates 3.5, i.e. the fundamental frequency Ω in units of the analysis subband frequency spacing $\Delta\omega$. In other words, the relative distance, i.e. the distance on the frequency axis divided by the analysis subband frequency spacing $\Delta\omega$, between the two analysis subbands contributing to the synthesis subband which is to be generated, should best approximate the relative fundamental frequency, i.e. the fundamental frequency Ω divided by the analysis subband frequency spacing $\Delta\omega$. This is also expressed by formulas (11) and leads to the choice $p_1 = 1, p_2 = 2$.

As shown in Fig. 17, the synthesis subband with index 8, i.e. reference sign 1710, is obtained from a cross product formed from the analysis subbands with index $(n - p_1) = 8 - 1 = 7$, i.e. reference sign 1706, and $(n + p_2) = 8 + 2 = 10$, i.e. reference sign 1708. For the synthesis subband with index 9, a cross product is formed from analysis subbands with index $(n - p_1) = 9 - 1 = 8$, i.e. reference sign 1707, and $(n + p_2) = 9 + 2 = 11$, i.e. reference sign 1709. This process of forming cross products is symbolized by the diagonal dashed/dotted arrow pairs, i.e. arrow pair 1712, 1713 and 1714, 1715, respectively. It can be seen from Fig. 17 that the partial frequency 7Ω is positioned more prominently in subband 1710 than in subband 1711. Consequently, it is to be expected

that for realistic filter responses, there will be more cross terms around synthesis subband with index 8, i.e. subband 1710, which add beneficially to the synthesis of a high quality sinusoid at frequency $(T-r)\omega+r(\omega+\Omega)=T\omega+r\Omega=6\Omega+\Omega=7\Omega$.

Fig. 18 illustrates a possible implementation of an additional cross term processing step for $r=2$ in the modulated filterbank of Fig. 16 which leads to the regeneration of the partial frequency at 8Ω. The index shifts (p_1,p_2) may be selected as a multiple of $(r,T-r)=(2,1)$, such that p_1+p_2 approximates 3.5, i.e. the fundamental frequency Ω in units of the analysis subband frequency spacing $\Delta\omega$. This leads to the choice $p_1=2, p_2=1$. As shown in Fig. 18, the synthesis subband with index 9, i.e. reference sign 1810, is obtained from a cross product formed from the analysis subbands with index $(n-p_1)=9-2=7$, i.e. reference sign 1806, and $(n+p_2)=9+1=10$, i.e. reference sign 1808. For the synthesis subband with index 10, a cross product is formed from analysis subbands with index $(n-p_1)=10-2=8$, i.e. reference sign 1807, and $(n+p_2)=10+1=11$, i.e. reference sign 1809. This process of forming cross products is symbolized by the diagonal dashed/dotted arrow pairs, i.e. arrow pair 1812, 1813 and 1814, 1815, respectively. It can be seen from Fig. 18 that the partial frequency 8Ω is positioned slightly more prominently in subband 1810 than in subband 1811. Consequently, it is to be expected that for realistic filter responses, there will be more direct and/or cross terms around synthesis subband with index 9, i.e. subband 1810, which add beneficially to the synthesis of a high quality sinusoid at frequency $(T-r)\omega+r(\omega+\Omega)=T\omega+r\Omega=2\Omega+6\Omega=8\Omega$.

In the following, reference is made to Figures 23 and 24 which illustrate the Max-Min optimization based selection procedure (12) for the index shift pair (p_1,p_2) and r according to this rule for $T=3$. The chosen target subband. index is $n=18$ and the top diagram furnishes an example of the magnitude of a subband signal for a given time index. The list of positive integers is given here by the seven values $L=\{2,3,...,8\}$.

Fig. 23 illustrates the search for candidates with $r=1$. The target or synthesis subband is shown with the index $n=18$. The dotted line 2301 highlights the subband with the index $n=18$ in the upper analysis subband range and the lower synthesis subband range. The

possible index shift pairs are $(p_1, p_2) = \{(2,4),(3,6),\ldots,(8,16)\}$, for $l = 2,3,\ldots,8$, respectively, and the corresponding analysis subband magnitude sample index pairs, i.e. the list of subband index pairs that are considered for determining the optimal cross term, are $\{(16,22),(15,24),\ldots,(10,34)\}$. The set of arrows illustrate the pairs under consideration. As an example, the pair $(15,24)$ denoted by the reference signs 2302 and 2303 is shown. Evaluating the minimum of these magnitude pairs gives the list $(0,4,1,0,0,0,0)$ of respective minimum magnitudes for the possible list of cross terms. Since the second entry for $l = 3$ is maximal, the pair $(15,24)$ wins among the candidates with $r = 1$, and this selection is depicted by the thick arrows.

Fig. 24 similarly illustrates the search for candidates with $r = 2$. The target or synthesis subband is shown with the index $n = 18$. The dotted line 2401 highlights the subband with the index $n = 18$ in the upper analysis subband range and the lower synthesis subband range. In this case, the possible index shift pairs are $(p_1, p_2) = \{(4,2),(6,3),\ldots,(16,8)\}$ and the corresponding analysis subband magnitude sample index pairs are $\{(14,20),(12,21),\ldots,(2,26)\}$, of which the pair (6,24) is represented by the reference signs 2402 and 2403. Evaluating the minimum of these magnitude pairs gives the list $(0,0,0,0,3,1,0)$. Since the fifth entry is maximal, i.e. $l = 6$, the pair $(6,24)$ wins among the candidates with $r = 2$, as depicted by the thick arrows. Overall, since the minimum of the corresponding magnitude pair is smaller than that of the selected subband pair for $r = 1$, the final selection for target subband index $n = 18$ falls on the pair $(15,24)$ and $r = 1$.

It should further more be noted that when the input signal $z(t)$ is a harmonic series with a fundamental frequency Ω , i.e. with a fundamental frequency which corresponds to the cross product enhancement pitch parameter, and Ω is sufficiently large compared to the frequency resolution of the analysis filter bank, the analysis subband signals $x_n(k)$ given by formula (6) and $x'_n(k)$ given by formula (8) are good approximations of the analysis of the input signal $z(t)$ where the approximation is valid in different subband regions. It follows from a comparison of the formulas (6) and (8-10) that a harmonic phase evolution along the frequency axis of the input signal $z(t)$ will be extrapolated correctly by the present invention. This holds in particular for a pure pulse train. For the output audio

quality, this is an attractive feature for signals of pulse train like character, such as those produced by human voices and some musical instruments.

Figures 25, 26 and 27 illustrate the performance of an exemplary implementation of the inventive transposition for a harmonic signal in the case $T=3$. The signal has a fundamental frequency 282.35 Hz and its magnitude spectrum in the considered target range of 10 to 15 kHz is depicted in Fig. 25. A filter bank of $N=512$ subbands is used at a sampling frequency of 48 kHz to implement the transpositions. The magnitude spectrum of the output of a third order direct transposer (T=3) is depicted in Fig 26. As can be seen, every third harmonic is reproduced with high fidelity as predicted by the theory outlined above, and the perceived pitch will be 847 Hz, three times the original one. Fig. 27 shows the output of a transposer applying cross term products. All harmonics have been recreated up to imperfections due to the approximative aspects of the theory. For this case, the side lobes are about 40 dB below the signal level and this is more than sufficient for regeneration of high frequency content which is perceptually indistinguishable from the original harmonic signal.

In the following, reference is made to Fig. 28 and Fig. 29 which illustrate an exemplary encoder 2800 and an exemplary decoder 2900, respectively, for unified speech and audio coding (USAC). The general structure of the USAC encoder 2800 and decoder 2900 is described as follows: First there may be a common pre/postprocessing consisting of an MPEG Surround (MPEGS) functional unit to handle stereo or multi-channel processing and an enhanced SBR (eSBR) unit 2801 and 2901, respectively, which handles the parametric representation of the higher audio frequencies in the input signal and which may make use of the harmonic transposition methods outlined in the present document. Then there are two branches, one consisting of a modified Advanced Audio Coding (AAC) tool path and the other consisting of a linear prediction coding (LP or LPC domain) based path, which in turn features either a frequency domain representation or a time domain representation of the LPC residual. All transmitted spectra for both, AAC and LPC, may be represented in MDCT domain following quantization and arithmetic coding. The time domain representation uses an ACELP excitation coding scheme.

The enhanced Spectral Band Replication (eSBR) unit 2801 of the encoder 2800 may comprise the high frequency reconstruction systems outlined in the present document. In

particular, the eSBR unit 2801 may comprise an analysis filter bank 301 in order to generate a plurality of analysis subband signals. This analysis subband signals may then be transposed in a non-linear processing unit 302 to generate a plurality of synthesis subband signals, which may then be inputted to a synthsis filter bank 303 in order to generate a high frequency component. In the eSBR unit 2801, on the encoding side, a set of information may be determined on how to generate a high frequency component from the low frequency component which best matches the high frequency component of the original signal. This set of information may comprise information on signal characteristics, such as a predominant fundamental frequency Ω, on the spectral envelope of the high frequency component, and it may comprise information on how to best combine analysis subband signals, i.e. information such as a limited set of index shift pairs (p_1,p_2). Encoded data related to this set of information is merged with the other encoded information in a bitstream multiplexer and forwarded as an encoded audio stream to a corresponding decoder 2900.

The decoder 2900 shown in Fig. 29 also comprises an enhanced Spectral Bandwidth Replication (eSBR) unit 2901. This eSBR unit 2901 receives the encoded audio bitstream or the encoded signal from the encoder 2800 and uses the methods outlined in the present document to generate a high frequency component of the signal, which is merged with the decoded low frequency component to yield a decoded signal. The eSBR unit 2901 may comprise the different components outlined in the present document. In particular, it may comprise an analysis filter bank 301, a non-linear processing unit 302 and a synthesis filter bank 303. The eSBR unit 2901 may use information on the high frequency component provided by the encoder 2800 in order to perform the high frequency reconstruction. Such information may be a fundamental frequency Ω of the signal, the spectral envelope of the original high frequency component and/or information on the analysis subbands which are to be used in order to generate the synthesis subband signals and ultimately the high frequency component of the decoded signal.

Furthermore, Figs. 28 and 29 illustrate possible additional components of a USAC encoder/decoder, such as:

- a bitstream payload demultiplexer tool, which separates the bitstream payload into the parts for each tool, and provides each of the tools with the bitstream payload information related to that tool;
- a scalefactor noiseless decoding tool, which takes information from the bitstream payload demultiplexer, parses that information, and decodes the Huffman and DPCM coded scalefactors;
- a spectral noiseless decoding tool, which takes information from the bitstream payload demultiplexer, parses that information, decodes the arithmetically coded data, and reconstructs the quantized spectra;
- an inverse quantizer tool, which takes the quantized values for the spectra, and converts the integer values to the non-scaled, reconstructed spectra; this quantizer is preferably a companding quantizer, whose companding factor depends on the chosen core coding mode;
- a noise filling tool, which is used to fill spectral gaps in the decoded spectra, which occur when spectral values are quantized to zero e.g. due to a strong restriction on bit demand in the encoder;
- a rescaling tool, which converts the integer representation of the scalefactors to the actual values, and multiplies the un-scaled inversely quantized spectra by the relevant scalefactors;
- a M/S tool, as described in ISO/IEC 14496-3;
- a temporal noise shaping (TNS) tool, as described in ISO/IEC 14496-3;
- a filter bank / block switching tool, which applies the inverse of the frequency mapping that was carried out in the encoder; an inverse modified discrete cosine transform (IMDCT) is preferably used for the filter bank tool;
- a time-warped filter bank / block switching tool, which replaces the normal filter bank / block switching tool when the time warping mode is enabled; the filter bank preferably is the same (IMDCT) as for the normal filter bank, additionally the windowed time domain samples are mapped from the warped time domain to the linear time domain by time-varying resampling;
- an MPEG Surround (MPEGS) tool, which produces multiple signals from one or more input signals by applying a sophisticated upmix procedure to the input signal(s) controlled by appropriate spatial parameters; in the USAC context, MPEGS is preferably used for coding a multichannel signal, by transmitting parametric side information alongside a transmitted downmixed signal;

- a Signal Classifier tool, which analyses the original input signal and generates from it control information which triggers the selection of the different coding modes; the analysis of the input signal is typically implementation dependent and will try to choose the optimal core coding mode for a given input signal frame; the output of the signal classifier may optionally also be used to influence the behaviour of other tools, for example MPEG Surround, enhanced SBR, time-warped filterbank and others;
- a LPC filter tool, which produces a time domain signal from an excitation domain signal by filtering the reconstructed excitation signal through a linear prediction synthesis filter; and
- an ACELP tool, which provides a way to efficiently represent a time domain excitation signal by combining a long term predictor (adaptive codeword) with a pulse-like sequence (innovation codeword).

Fig. 30 illustrates an embodiment of the eSBR units shown in Figs. 28 and 29. The eSBR unit 3000 will be described in the following in the context of a decoder, where the input to the eSBR unit 3000 is the low frequency component, also known as the lowband, of a signal and possible additional information regarding specific signal characteristics, such as a fundamental frequency Ω, and/or possible index shift values (p_1,p_2). On the encoder side, the input to the eSBR unit will typically be the complete signal, whereas the output will be additional information regarding the signal characteristics and/or index shift values.

In Fig. 30 the low frequency component 3013 is fed into a QMF filter bank, in order to generate QMF frequency bands. These QMF frequency bands are not be mistaken with the analysis subbands outlined in this document. The QMF frequency bands are used for the purpose of manipulating and merging the low and high frequency component of the signal in the frequency domain, rather than in the time domain. The low frequency component 3014 is fed into the transposition unit 3004 which corresponds to the systems for high frequency reconstruction outlined in the present document. The transposition unit 3004 may also receive additional information 3011, such as the fundamental frequency Ω of the encoded signal and/or possible index shift pairs (p_1,p_2) for subband selection. The transposition unit 3004 generates a high frequency component 3012, also known as highband, of the signal, which is transformed into the

frequency domain by a QMF filter bank 3003. Both, the QMF transformed low frequency component and the QMF transformed high frequency component are fed into a manipulation and merging unit 3005. This unit 3005 may perform an envelope adjustment of the high frequency component and combines the adjusted high frequency component and the low frequency component. The combined output signal is re-transformed into the time domain by an inverse QMF filter bank 3001.

Typically the QMF filter banks comprise 64 QMF frequency bands. It should be noted, however, that it may be beneficial to down-sample the low frequency component 3013, such that the QMF filter bank 3002 only requires 32 QMF frequency bands. In such cases, the low frequency component 3013 has a bandwidth of $f_s/4$, where f_s is the sampling frequency of the signal. On the other hand, the high frequency component 3012 has a bandwidth of $f_s/2$.

The method and system described in the present document may be implemented as software, firmware and/or hardware. Certain components may e.g. be implemented as software running on a digital signal processor or microprocessor. Other component may e.g. be implemented as hardware and or as application specific integrated circuits. The signals encountered in the described methods and systems may be stored on media such as random access memory or optical storage media. They may be transferred via networks, such as radio networks, satellite networks, wireless networks or wireline networks, e.g. the internet. Typical devices making use of the method and system described in the present document are set-top boxes or other customer premises equipment which decode audio signals. On the encoding side, the method and system may be used in broadcasting stations, e.g. in video headend systems.

The present document outlined a method and a system for performing high frequency reconstruction of a signal based on the low frequency component of that signal. By using combinations of subbands from the low frequency component, the method and system allow the reconstruction of frequencies and frequency bands which may not be generated by transposition methods known from the art. Furthermore, the described HTR method and system allow the use of low cross over frequencies and/or the generation of large high frequency bands from narrow low frequency bands.

CLAIMS

1. A system for generating a high frequency component of a signal from a low frequency component of the signal, comprising:
   - an analysis filter bank (301) providing a plurality of analysis subband signals of the low frequency component of the signal;
   - a non-linear processing unit (302) to generate a synthesis subband signal with a synthesis frequency by modifying the phase of a first and a second of the plurality of analysis subband signals and by combining the phase-modified analysis subband signals; and
   - a synthesis filter bank (303) for generating the high frequency component of the signal from the synthesis subband signal.

2. The system according to claim 1, wherein the non-linear processing unit (302) comprises:
   - a multiple-input-single-output unit (800-n) of a first and second transposition order generating the synthesis subband signal (803) from the first (801) and the second (802) analysis subband signals with a first and a second analysis frequency, respectively, wherein
   - the first analysis subband signal (801) is phase-modified by the first transposition order;
   - the second analysis subband signal (803) is phase-modified by the second transposition order; and
   - the synthesis frequency corresponds to the first analysis frequency multiplied by the first transposition order plus the second analysis frequency multiplied by the second transposition order.

3. The system according to claim 2, wherein
   - the phase modification is a phase multiplication with a transposition order;
   - the first analysis frequency is ω;

- the second analysis frequency is $(\omega+\Omega)$
- the first transposition order is (T-r);
- the second transposition order is r;
- T>1; and
- $1 \leq r < T$;

such that the synthesis frequency is $(T-r)\cdot\omega + r\cdot(\omega+\Omega)$.

4. The system according to claim 1, further comprising:
   - a gain unit (902) for multiplying the synthesis subband signal (803) by a gain parameter.
   -
5. The system according to claim 2, further comprising
   - a plurality of multiple-input-single-output units (800-n) and/or a plurality of non-linear processing units which generate a plurality of partial synthesis subband signals (803) with the synthesis frequency; and
   - a subband summing unit (702) for combining the plurality of partial synthesis subband signals.

6. The system according to claim 2, wherein the non-linear processing unit (302) further comprises:
   - a direct processing unit (401) for generating a further synthesis subband signal from a third of the plurality of analysis subband signals; and
   - a subband summing unit for combining synthesis subband signals with the synthesis frequency.

7. The system according to claim 2, wherein
   - the subband summing unit ignores the synthesis subband signals generated in the multiple-input-single-output units (800-n) if the minimum of the magnitude of the first (801) and second (802) analysis subband signals is smaller than a pre-defined fraction of the magnitude of the signal.

8. The system according to claim 6, wherein the direct processing unit (401) comprises:
   - a single-input-single-output unit (401-n) of a third transposition order T', generating the synthesis subband signal from the third analysis subband signal exhibiting a third analysis frequency, wherein
   - the third analysis subband signal is phase-modified by the third transposition order T';
   - T' is greater than one; and
   - the synthesis frequency corresponds to the third analysis frequency multiplied by the third transposition order.

9. The system according to claim 1, wherein
   - the signal comprises a fundamental frequency; and
   - the analysis filter bank (301) exhibits a frequency spacing which is associated with the fundamental frequency of the signal.

10. The system according to claim 3, wherein
   - the analysis filter bank (301) has N analysis subbands at an essentially constant subband spacing of $\Delta\omega$;
   - an analysis subband is associated with an analysis subband index n, with $n \in \{1,...,N\}$;
   - the synthesis filter bank (303) has a synthesis subband;
   - the synthesis subband is associated with a synthesis subband index n; and
   - the synthesis subband and the analysis subband with index n each comprise frequency ranges which relate to each other through the factor T.

11. The system according to claim 10, wherein
   - the synthesis subband signal (803) is associated with the synthesis subband with index n;
   - the first analysis subband signal (801) is associated with an analysis subband with index $n-p_1$;
   - the second analysis subband signal (802) is associated with an analysis subband with index $n+p_2$; and

– the system further comprises an index selection unit for selecting p_1 and p_2.

12. The system according to claim 11, wherein

– the index selection unit is operable to select the index shifts p_1 and p_2 from a limited list of pairs (p_1, p_2) stored in an index storing unit.

13. The system according to claim 12, wherein

– the index selection unit is operable to select the pair (p_1, p_2) such that the minimum value of a set comprising the magnitude of the first analysis subband signal and the magnitude of the second analysis subband signal is maximized.

14. The system according to claim 11, wherein the index selection unit is operable to determine a limited list of pairs (p_1, p_2) such that

– the index shift $p_1 = r \cdot l$;

– the index shift $p_2 = (T-r) \cdot l$; and

– l is a positive integer.

15. The system according to claim 14, wherein

– the index selection unit is operable to select the parameters l and r such that the minimum value of the set comprising the magnitude of the first analysis subband signal and the magnitude of the second analysis subband signal is maximized.

16. The system according to claim 11, wherein the index selection unit is operable to select the index shifts p_1 and p_2 based on a characteristic of the signal.

17. The system according to claim 16, wherein

– the signal comprises a fundamental frequency Ω;

– the index selection unit is operable to select the index shifts p_1 and p_2 such that

  – their sum of the index shifts p_1+p_2 approximates the fraction $\Omega/\Delta\omega$; and

  – their fraction p_1/p_2 is a multiple of $r/(T-r)$.

18. The system according to claim 16, wherein

- the signal comprises a fundamental frequency Ω;
- the index selection unit is operable to select the index shifts p_1 and p_2 such that
  - their sum of the index shifts p_1+p_2 approximates the fraction $\Omega/\Delta\omega$; and
  - the fraction p_1/p_2 equals $r/(T-r)$.

19. The system according to claim 1, further comprising:

- an analysis window (2001), which isolates a pre-defined time interval of the low frequency component around a pre-defined time instance k; and
- a synthesis window (2201), which isolates a pre-defined time interval of the high frequency component around the pre-defined time instance k.

20. The system according to claim 19, wherein

- the synthesis window (2201) is a time-scaled version of the analysis window (2001).

21. A system for decoding a signal comprising:

- a transposition unit (102) according to claim 1 for generating the high frequency component of the signal from the low frequency component of the signal.

22. The system according to claim 21, wherein the signal is a speech and/or audio signal:

23. The system according to claim 21, further comprising:

- a core decoder (101) for decoding the low frequency component of the signal.

24. The system according to claim 21, further comprising:

- an upsampler (104) for performing an upsampling of the low frequency component to yield an upsampled low frequency component;
- an envelope adjuster (103) to shape the high frequency component; and
- a component summing unit to determine the decoded signal as the sum of the upsampled low frequency component and the adjusted high frequency component.

25. The system according to claim 21, further comprising

- a subband selection reception unit for receiving information which allows the selection of the first (801) and second (802) analysis subband signals from which the synthesis subband signal (803) is to be generated.

26. The system according to claim 25, wherein

- the information is associated with a fundamental frequency Ω of the signal.

27. The system according to claim 25, wherein

- the information comprises a list of pairs of first (801) and second (802) analysis subband signals.

28. The system according to claim 24, further comprising:

- an envelope reception unit for receiving information related to the envelope of the high frequency component of the signal.

29. The system according to claim 23, further comprising:

- an input unit for receiving the signal, comprising the low frequency component; and
- an output unit for providing the decoded signal, comprising the low and the generated high frequency component.

30. An encoded signal, comprising:

- information related to a low frequency component of the decoded signal, wherein the low frequency component comprises a plurality of analysis subband signals;
- information related to which two of the plurality of analysis subband signals are to be selected to generate a high frequency component of the decoded signal by transposing the selected two analysis subband signals.

31. A system for encoding a signal, comprising:

- a splitting unit for splitting the signal into a low frequency component and into a high frequency component;
- a core encoder for encoding the low frequency component;

- a frequency determination unit for determining a fundamental frequency Ω of the signal; and
- a parameter encoder for encoding the fundamental frequency Ω, wherein the fundamental frequency Ω is used to regenerate the high frequency component of the signal.

32. The system according to claim 31, further comprising:

- an envelope determination unit for determining the spectral envelope of the high frequency component; and
- an envelope encoder for encoding the spectral envelope.

33. A system for encoding a signal, comprising:

- a splitting unit for splitting the signal into a low frequency component and into a high frequency component;
- a core encoder for encoding the low frequency component;
- an analysis filter bank providing a plurality of analysis subband signals of the low frequency component of the signal;
- a subband pair determination unit for determining a first and a second subband signal for generating a high frequency component of the signal; and
- an index encoder for encoding index numbers representing the first and the second subband signal.

34. A method for performing high frequency reconstruction of a high frequency component from a low frequency component of a signal, comprising:

- providing (301) a first subband signal of the low frequency component from a first frequency band and a second subband signal of the low frequency component from a second frequency band;
- transposing (302) the first and the second subband signal by a first and a second transposition factor, respectively;
- combining (303) the transposed first and second subband signals to yield a high frequency component from a high frequency band.

35. The method according to claim 34, wherein
- the high frequency band corresponds to the sum of the first frequency band multiplied by the first transposition factor and the second frequency band multiplied by the second transposition factor.

36. The method according to claim 34, wherein the transposing step comprises:
- multiplying the first frequency band of the first subband signal with the first transposition factor; and
- multiplying the second frequency band of the second subband signal with the second transposition factor.

37. The method according to claim 34, wherein the providing step comprises:
- filtering the low frequency component by an analysis filter bank (301) to generate a first and a second subband signal.

38. The method according to claim 34, wherein the combining step comprises:
- multiplying the first and the second transposed subband signals to yield a high subband signal; and
- inputting the high subband signal into a synthesis filter bank to generate the high frequency component.

39. A method for decoding an encoded signal, wherein the encoded signal
- is derived from an original signal; and
- represents only a portion of frequency subbands of the original signal below a cross-over frequency (1005);

wherein the method comprises
- providing (301) a first and a second frequency subband of the encoded signal;
- transposing (302) the frequency subbands by a first transposition factor and a second transposition factor, respectively; and
- generating (303) a high frequency subband from the first and second transposed frequency subbands, wherein the high frequency subband is above the cross-over frequency band.

40. The method according to claim 39, wherein
- the high frequency subband corresponds to the sum of the first frequency subband multiplied by the first transposition factor and the second frequency subband multiplied by the second transposition factor.

41. The method according to claim 39, wherein the transposing step comprises:
- performing a phase multiplication of the signal in the first frequency subband with the first transposition factor; and
- performing a phase multiplication of the signal in the second frequency subband with the second transposition factor.

42. A method for encoding a signal, comprising:
- filtering the signal to isolate a low frequency of the signal;
- encoding the low frequency component of the signal;
- providing a plurality of analysis subband signals of the low frequency component of the signal;
- determining a first and a second subband signal for generating a high frequency component of the signal; and
- encoding information representing the first and the second subband signal.

43. A set-top box for decoding a received multimedia signal comprising an audio signal, the set-top box comprising:
- a transposition unit (102) according to claim 1 for generating the high frequency component of the signal from the low frequency component of the audio signal.

44. A software program adapted for execution on a processor and for performing the method steps of claim 34 when carried out on a computing device.

45. A storage medium comprising a software program adapted for execution on a processor and for performing the method steps of claim 34 when carried out on a computing device.

46. A computer program product comprising executable instructions for performing the method of claim 34 when executed on a computer.

47. The system according to claim 23, wherein the core decoder (101) is based on a coding scheme being one of: Dolby E, Dolby Digital, AAC.