



(43) International Publication Date
29 November 2012 (29.11.2012)

(51) International Patent Classification:

G10L 19/08 (2006.01) G10L 19/00 (2006.01)
G10L 19/02 (2006.01)

(21) International Application Number:

PCT/US2011/037336

(22) International Filing Date:

20 May 2011 (20.05.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant (for all designated States except US):
GOOGLE INC. [US/US]; 1600 Amphitheatre Parkway,
Mountain View, California 94043 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ZAKIZADEH SHABESTARY, Turaj** [IR/US]; 2200 Pacific Ave, Apt 4B, San Francisco, California 94115 (US). **SKOGLUND, Jan** [US/US]; 235 Noe Street, San Francisco, California 94114 (US).

(74) Agent: **CAMMARATA, Michael R.**; c/o Birch, Stewart, Kolasch & Birch, LLP, P.O. Box 747, Falls Church, Virginia 22040-0747 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- of inventorship (Rule 4.17(iv))

Published:

- with international search report (Art. 21(3))
- with amended claims (Art. 19(1))

(54) Title: REDUNDANT CODING UNIT FOR AUDIO CODEC

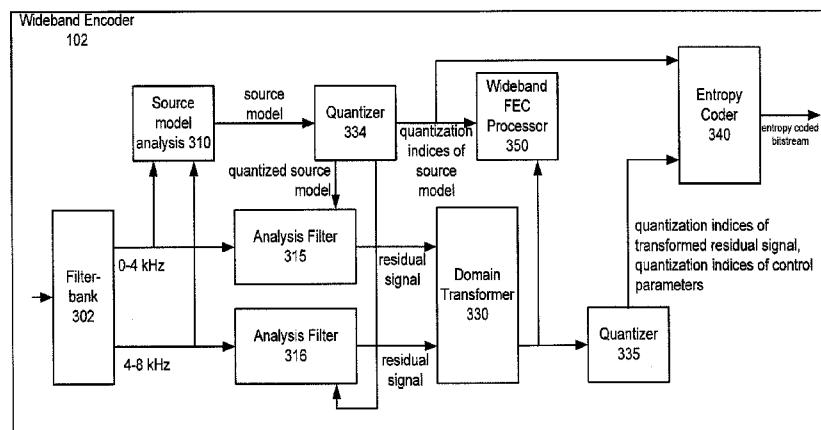


FIG. 3

(57) Abstract: A method of generating multiple payloads encoded at different coding rates for inclusion in a data packet transmitted across a packet data network includes extracting from a source model representing a spectral envelope of the input signal and a residual signal by filtering input signal with filter coefficients derived from the source model to remove effects of pitch from the first frequency band signal. The source model and the residual model are stored. When payload is to be encoded at a lower coding rate, the residual model is scaled by a scaling factor and quantized. The scaled residual model and the source model may also be entropy coded to generate a redundant bitstream. The redundant bitstream is used to generate payloads for packets transmitted over a network.

WO 2012/161675 A1

REDUNDANT CODING UNIT FOR AUDIO CODEC

TECHNICAL FIELD

[0001] The technical field relates to packet loss concealment in communication systems (such as Voice over IP, also referred to as VoIP), having an audio codec (coder/decoder). One such codec may be iSAC.

BACKGROUND

[0002] Telephone communication originally relied on dedicated connections between callers. Thus, every ongoing telephone conversation required a physical, real-time, connection to enable real-time communication. Real-time communication refers to communication where the delay between one user speaking and another user hearing the speech is so short that it is imperceptible or nearly imperceptible. In recent years, advances in communication technology have allowed packet-switched networks, such as the Internet, to support real-time communication.

[0003] VoIP is one audio communication approach enabling real-time communication over packet-switched networks. Instead of a dedicated connection between callers, an audio signal is broken up into short time segments by an audio coder, and the time segments are transmitted individually as audio frames in packets. The packets are received by the receiver, the audio frames are extracted, and the short time segments are reassembled by an audio decoder into the original audio signal, enabling the receiver to hear the transmitted audio signal.

[0004] Real time audio communication over packet-switched networks has brought with it unique challenges. The available bandwidth of the network may be limited, and may change over time. Packets may also get lost or corrupted. A packet is considered lost, when it fails to arrive at the intended receiver within some amount of time, even if the packet does eventually arrive at the receiver.

[0005] One approach for dealing with lost packets is Backward Error Correction (BEC), where the receiver notifies the transmitter that an expected packet was not received, causing the transmitter to re-transmit the expected packet. While viable for tasks such as file

transmission, BEC is not desirable for a real-time communication system. In real-time audio communication re-transmission is not a viable option because it typically results in a large delay before the missing packet is received by the receiver. Waiting for re-transmission of a packet would result in the loss of the real-time nature of the communication.

[0006] Another approach for addressing the problem of lost packets is Forward Error Correction (FEC). In FEC the transmitter of the audio data adds redundant audio data to the packets as they are generated. Specifically, a packet may contain audio data (audio frame) corresponding to a time period t_2 and the immediately preceding time period t_1 . The second packet may contain audio data corresponding to the time period t_3 and the immediately preceding time period t_2 . The third packet may contain audio data corresponding to time period t_4 and the immediately preceding time period t_3 . If the second packet is lost, it is possible to recreate the full audio segment of t_1 , t_2 , and t_3 , from only the first packet and the third packet, because the third packet contains audio data corresponding to time period t_2 .

[0007] RFC 2198 (Internet Engineering Task Force Request for Comments 2198) describes an FEC approach, where a previously encoded and transmitted audio frame is aggregated with the current frame and is re-transmitted together with the current frame. In the simplest case packet n contains the payload of frame $n-1$ (redundant payload), and the payload of frame n (main payload). Therefore, in this example, it is necessary to have twice the transmission rate to maintain the same data throughput as compared to when no redundant data is transmitted. In other word, the effective bandwidth available for communication of data is halved, because redundant data is being transmitted along with primary data. Although this scheme is simple to implement and low-complexity to operate, it is expensive in terms of required bandwidth.

[0008] RFC 2198 describes another approach where the redundant payload is encoded with a different encoder than the primary payload. However, that approach requires two coders to be executed by the transmitter, and two decoders to be operated by the receiver. A coder and decoder normally run continuously, with associated memory buffering of incoming/outgoing data, thus the approach is expensive in terms of processing and memory load, and is impractical for situations where the processing power and memory are expensive, or altogether unavailable, such as on mobile communication devices.

[0009] The present invention recognizes the problem posed by lost packets in real-time audio communication over packet switched networks, and provides a solution that avoids the disadvantages of the above examples.

SUMMARY

[0010] In an embodiment, a method of generating multiple payloads encoded at different coding rates for inclusion in a data packet transmitted across a packet data network, the data packet containing an encoded audio source signal, may include separating the audio source signal into a first frequency band signal and a second frequency band signal, the first frequency band being lower than the second frequency band. The method may also include extracting from the first frequency band signal a source model representing linear dependencies of the first frequency band signal, generating a residual signal by filtering the first frequency band signal with a filter having filter coefficients derived from the source model to remove short-term and long-term linear dependencies from the first frequency band signal, transforming the residual signal into a transform domain, scaling the transformed residual signal with a first scale factor, quantizing the scaled transformed residual signal, quantizing the source model to create quantization indices of the source model, entropy encoding the quantization indices of the scaled transformed residual signal and the quantization indices of the source model to generate a redundant bitstream and constructing the payloads from the redundant bitstream.

[0011] In an embodiment, the method may also include extracting from the second frequency band signal a second source model representing linear dependencies of the second frequency band signal, generating a second residual signal by filtering the second frequency band signal with filter coefficients derived from the second source model to remove linear dependencies from the second frequency band signal, transforming the second residual signal into the transform domain, scaling the transformed second residual signal with a second scale factor, quantizing the scaled transformed second residual signal to create quantization indices of the transformed residual signal, quantizing the second source model to create quantization indices of the source model, entropy encoding the quantization indices of the scaled transformed second residual signal and the quantization indices of the second source model to generate a second redundant bitstream, and constructing the payloads from the second redundant bitstream.

[0012] In an embodiment, the method may also include storing in a memory the transformed residual signal and the quantization indices of the source model prior to the scaling of the transformed residual signal, and extracting from the memory the stored transformed residual signal and the quantization indices of the source model when the redundant payload is to be encoded at a lower coding rate than a coding rate of a primary payload.

[0013] In an embodiment, the step of separating the audio source signal into the first frequency band signal and the second frequency band signal may include dividing the first frequency band into a first sub-band and a second sub-band, the first sub-band being lower than the second sub-band.

[0014] In an embodiment, the method may further include generating a data frame including an encoded segment of the audio source signal corresponding to a first time period, and an encoded segment of the audio source signal corresponding to a second time period different than the first time period, the second time period preceding the first time period.

[0015] In an embodiment, the segment of the audio source signal corresponding to the first time period is encoded at a higher coding rate than the coding rate of the segment of audio source data corresponding to the second time period.

[0016] In an embodiment, the first scale factor and the second scale factor are set independently of each other, based on the coding rate of the segment of audio source data corresponding to the second time period.

[0017] In an embodiment, first scale factor and the second scale factor are each independently set to a value between 0.4 and 0.5, inclusive, to reach a suitable overall-quality vs concealment-quality trade off, at packet loss rates in the range of 10% to 15%, inclusive.

[0018] In an embodiment, the method may also include downsampling the second frequency band signal prior to the extracting of the second source model.

[0019] In an embodiment, an encoding apparatus for encoding a source audio signal at different coding rates to generate multiple payloads included in data packets transmitted across a packet data network, includes a filter-bank configured to separate the audio source signal into a first frequency band signal and a second frequency band signal, the first frequency band being lower than the second frequency band. The apparatus may also include a source model analysis unit configured to generate a source model representing linear dependencies of the first frequency band signal, an analysis filter having its filter coefficients

derived from the source model and configured to filter the first frequency band signal to generate a residual signal, a domain transformer transforming the residual signal into a transform domain, a multiplier multiplying the transformed residual signal with a first scale factor, a quantizer quantizing the scaled transformed residual signal, and quantizing the source model to create associated quantization indices of the source model and the scaled transformed residual signal, and an entropy coder encoding the quantization indices of the scaled transformed residual signal and the quantization indices of the source model to generate a redundant bitstream.

[0020] In an embodiment, the apparatus may also include a second source model analysis unit configured to extract from the second frequency band signal a second source model representing linear dependencies of the second frequency band signal, a second analysis filter having filter coefficients derived from the second source model and configured to filter the second frequency band signal to generate a second residual signal, a second domain transformer transforming the second residual signal into the transform domain, a second multiplier multiplying the transformed second residual signal with a second scale factor, a second quantizer quantizing the scaled transformed second residual signal, and quantizing the second source model to create quantization indices of the second source model and quantization indices of the scaled transformed second residual signal, and a second entropy coder encoding the quantization indices of the scaled transformed second residual signal and the quantization indices of the second source model to generate a second redundant bitstream.

[0021] In an embodiment, the apparatus may include a storage unit storing the transformed residual signal and the quantization indices of the source model prior to the multiplication by the multiplier, wherein the stored transformed residual signal and the quantization indices of the source model are extracted from the storage unit when the redundant payload is to be encoded at a lower coding rate than a coding rate of a primary payload.

[0022] In an embodiment, the filter bank is further configured to divide the first frequency band into a first sub-band and a second sub-band, the first sub-band being lower than the second sub-band.

[0023] In an embodiment, the apparatus may include a concatenation unit configured to generate a data frame including an encoded segment of the audio source signal corresponding to a first time period and an encoded segment of the audio source signal corresponding to a

second time period different than the first time period, the second time period preceding the first time period.

[0024] In an embodiment, the segment of the audio source signal corresponding to the first time period is encoded at a higher coding rate than the coding rate of the segment of audio source data corresponding to the second time period.

[0025] In an embodiment, the first scale factor and the second scale factor are set independently of each other, based on the coding rate of the segment of audio source data corresponding to the second time period.

[0026] In an embodiment, the first scale factor and the second scale factor are each independently set to a value between 0.4 and 0.5, inclusive, to reach a suitable overall-quality vs concealment-quality trade off, at packet loss rates in the range of 10% to 15%, inclusive.

[0027] In an embodiment, the apparatus may also include a downsampler configured to downsample the second frequency band signal prior to processing by the second source model analysis unit and the second analysis filter.

[0028] In an embodiment, a computer readable tangible recording medium is encoded with instructions, wherein the instructions when executed by a processor cause the processor to perform a method of generating multiple payloads encoded at different coding rates for inclusion in a data packet transmitted across a packet data network, the method including separating the audio source signal into a first frequency band signal and a second frequency band signal, the first frequency band being lower than the second frequency band. The method also includes extracting from the first frequency band signal a source model representing linear dependencies of the first frequency band signal, generating a residual signal by filtering the first frequency band signal with a filter having filter coefficients derived from the source model to remove short-term and long-term linear dependencies from the first frequency band signal, transforming the residual signal into a transform domain, scaling the transformed residual signal with a first scale factor, quantizing the scaled transformed residual signal, quantizing the source model to create quantization indices of the source model, entropy encoding the quantization indices of the scaled transformed residual signal and the quantization indices of the source model to generate a redundant bitstream, and constructing the payloads from the redundant bitstream.

[0029] In an embodiment, method performed by the processor may further include extracting from the second frequency band signal a second source model representing linear

dependencies of the second frequency band signal, generating a second residual signal by filtering the second frequency band signal with filter coefficients derived from the second source model to remove linear dependencies from the second frequency band signal, transforming the second residual signal into the transform domain, scaling the transformed second residual signal with a second scale factor, quantizing the scaled transformed second residual signal to create quantization indices of the transformed residual signal, quantizing the second source model to create quantization indices of the source model, entropy encoding the quantization indices of the scaled transformed second residual signal and the quantization indices of the second source model to generate a second redundant bitstream, and constructing the payloads from the second redundant bitstream.

BRIEF DESCRIPTION OF DRAWINGS

[0030] The present invention will become more fully understood from the detailed description given herein below and the accompanying drawings which are given by way of illustration only, and thus do not limit the present invention.

[0031] **FIG. 1** is a block diagram illustrating a communication system according to an embodiment of the present invention.

[0032] **FIG. 2** illustrates an example of the communication system of **FIG. 1** in greater detail.

[0033] **FIG. 3** illustrates an example of a wideband encoder according to an embodiment of the present invention.

[0034] **FIG. 4** illustrates an example of a wideband FEC processor according to an embodiment of the present invention.

[0035] **FIG. 5** illustrates an example of a super-wideband encoder according to an embodiment of the present invention.

[0036] **FIG. 6** illustrates an example of a super-wideband FEC processor according to an embodiment of the present invention.

[0037] **FIG. 7** illustrates an example of a process flow of the encoding process according to an embodiment of the present invention.

[0038] **FIG. 8** illustrates an example of a wideband decoder according to an embodiment of the present invention.

[0039] **FIG. 9** illustrates an example of a super-wideband decoder according to an embodiment of the present invention.

[0040] **FIG. 10** illustrates an example of a process flow of the decoding process according to an embodiment of the present invention.

[0041] **FIG. 11** illustrates an example of a computing device configured to perform encoding and decoding according to an embodiment of the present invention.

DETAILED DESCRIPTION

[0042] **Fig. 1** illustrates a communication system. Audio input is passed into one end of the system, and is ultimately output at the other end of the system. The communication can be concurrently bi-directional, as in a telephone conversation between two callers. The audio input can be generated by a user speaking, by a recording, or any other audio source. The audio input is supplied to encoding module **101**, where it is encoded and transmitted to packet network **104**.

[0043] Encoding module **101** encodes the audio input into multiple packets, which are transmitted over packet network/IP channel **104** to decoding module **109**. Packet network **104** can be any packet-switched network, whether using physical link connection and/or wireless link connections. Packet network **104** may also be a wireless communication network, and/or an optical link network. Packet network **104** conveys packets from encoding module **101** to decoding module **109**. Some of the packets sent may get lost.

[0044] Decoding module **109** receives packets conveyed by network **104** and decodes the packets into audio data.

[0045] **Fig. 2** illustrates additional details of the system of **Fig. 1**. The audio input may be sampled at a sampling frequency of 32 kHz or 16 kHz, as illustrated in **Fig. 2**. Audio sampled at 16 kHz corresponds to a bandwidth of 0-8kHz, and will be referred to as "wideband." Audio sampled at 32 kHz corresponds to a bandwidth of 0-16kHz. In this bandwidth, the frequency range 0-8kHz is referred as wideband, while the frequency range of 8-16kHz will be referred as "super-wideband." As can be appreciated, other frequency ranges could be selected, and the specific ranges noted are not limiting, but merely exemplary. For the purpose of an illustrative example, the frequency range 0-16 kHz (fs=32kHz) will be used below, thus including a wideband and a super-wideband range.

However, it is possible that only the wideband range will be present, without the super-wideband range.

[0046] Filter-bank **202** separates the incoming signal into the wideband signal and the super-wideband signal. The wideband signal is encoded by the wideband encoder **102**, while the super-wideband signal is encoded by super-wideband encoder **103**.

[0047] After the wideband and the super-wideband signals are encoded, the respective encoder produce encoded bitstreams which are concatenated and transmitted via an IP channel such as packet switched network **104**.

[0048] After transmission via the IP channel, the bitstream is received and separated into separate bitstreams for the wideband and the super-wideband signal, respectively. The wideband bitstream is decoded by wideband decoder **106**, while the super-wideband bitstream is decoded by super-wideband decoder **107**. Once the bitstreams are decoded by their respective decoders, the output signals are combined in the filter-bank **204**.

[0049] **Fig. 3** shows an example of an embodiment of wideband encoder **102**. Audio input is received in filter-bank **202**, where it is separated into a low band (0-4 kHz) and a high band (4-8 kHz), as illustrated in the figure.

[0050] Both the low band and the high band signals are analyzed by source model analysis **310**. Source model analysis **310** conducts source model analysis of the incoming audio signals and produces a corresponding source model for each of the low band and the high band signals.

[0051] In an embodiment, the source model may be derived by performing linear prediction coding (LPC) analysis together with pitch analysis on the incoming signals. A given frame of audio is described as a quasi-time-invariant linear filter, production filter, excited by a residual signal. The quasi-time-invariance is due to the fact that the production filter needs to be updated every 5 to 10 ms, therefore, within each sub-frame (5-10ms) the filter is time-invariant. The production filter captures the short-term and the long-term linear dependencies in the signal. Short-term dependencies may be modeled by LPC analysis, and the long-term dependencies may be modeled by pitch analysis. Translated to frequency domain, LPC analysis describes spectral envelope of the signal in question and pitch analysis reveals fine structure in the frequency domain.

[0052] The source model is passed from source model analysis **310** to quantizer **334**, where the source model is quantized. The process of quantization maps continuous values of a variable to a set of discrete values.

[0053] The encoding of the source model output from source model analysis **310** is rate independent. Thus, for all transmission rates the source model as generated by source modeling **301** is the same. The effect of available transmission rates is discussed with regard to wideband FEC processor **350**.

[0054] The output of the quantizer **334** is a quantized version of the source model, which is supplied as filter parameters of analysis filters **315** and **316**, and quantization indices of the source model, which are supplied to wideband FEC processor **350**. The analysis filter **315** takes as input the low band (0-4kHz) signal, and derives the residual signal of the low band signal.

[0055] The analysis filter **316** takes as input the high band (4-8kHz) signal and derives the residual signal of the high band signal.

[0056] Analysis filter **315** filters the digital signal input based on the quantized values of the source model. The analysis filter **315** can be implemented to perform the analysis in two steps - short-term analysis and long-term analysis. The short-term analysis filter is an all-zero filter, which can be implemented as a lattice filter, where filter coefficients are given by LPC analysis. The long-term analysis filter can be a pole-zero filter, with filter coefficients derived from the quantized pitch lag and gain (which are part of the source model). Long-term analysis filter is only applied to the low band signal as pitch structure very rarely extends beyond 4 kHz. The analysis filter **315** removes short-term and long-term structure (as determined during source modeling) from the input signal, outputting a residual signal.

[0057] The residual signal is transformed from the time domain into an alternate domain, for example the frequency domain through Discrete Fourier Transform (DFT), by domain transformer **330**. The domain transformer **330** can be implemented as an FFT (Fast Fourier Transform). Other transformations such as Modified Discrete Cosine Transform (MDCT) or Modified Lapped Transform (MLT) might be used instead of DFT. When the specification refers to an FFT and inverse FFT, it is to be understood that other transforms are contemplated and are not excluded by referring to the FFT.

[0058] The output of the domain transformer **330** is the residual signal in the transform domain. For example, the output may be DFT coefficients of the residual signal.

[0059] The DFT of the residual signal output from domain transformer **330**, along with quantization indices of the source model are supplied to wideband FEC processor **350**, where they are stored in FEC cabinet **410**. Additional details of FEC processor **350** are illustrated in **Fig. 4**.

[0060] The DFT of the residual model is also supplied to quantizer **335**, where the residual signal is quantized and a set of control parameters are derived and quantized. Then, quantization indices of the residual signal and quantization indices of control parameters are supplied to entropy coder **340**. The control parameters define a cumulative distribution function (CDF) of residual signal. In other words, the quantized control parameters can be defined as an Auto Regressive (AR) model for entropy coding of DFTs. Quantization indices of the source model are also supplied to entropy coder **340**.

[0061] The entropy coder **340** is progressive over the entire bit-stream, so to encode/decode index k , index $k-1$ had to be encoded/decoded. The entropy coder **340** may be implemented as a range encoder. The input to a range encoder is a sequence of indices with the associated CDF for each index. Indices are fed into the range encoder one by one. With each input, the state of the range encoder changes and when it reaches to a pre-defined state then a sequence of bits are generated and the state is modified accordingly. After the last index is inserted, depending on the state of the range encoder a sequence of bits is generated as a termination point. The order in which quantization indices are fed into the entropy coding is as follows: frame-size, bandwidth information, pitch lag, pitch gain, LP-shape, LP-gain, and residual signal.

[0062] The entropy coding of DFT coefficients is not any different than coding any other coefficients. The difference is in computation of CDFs. Other than DFTs, any other coefficient has a fixed CDF, which is used for entropy coding/decoding. An assumption is made that if DFT coefficients are properly normalized then there is a single CDF which describes the statistics of the normalized coefficients. For each DFT coefficient the normalization factor is the standard deviation of the coefficient in question. Given a set of DFT coefficients the spectral envelope is considered as an estimate for the standard deviation of each coefficient. Such an envelope is modeled by an Auto Regressive (AR) process. The AR coefficients can be computed by LPC analysis over the given DFT coefficients. The set of AR coefficients are referred by the "control parameters" which are mentioned previously.

U.S. Patent No. 7,756,350 describes additional aspects of using the AR process for selecting CDF's and is incorporated herein by reference in its entirety.

[0063] The output of entropy coder **340** is an entropy coded bitstream representing the wideband (0-8kHz) components of the incoming audio signal.

[0064] As illustrated in **Fig. 2**, the bitstream can then be concatenated with the bitstream generated by the super-wideband encoder to create packets for transmission over the IP channel.

[0065] **Fig. 4** illustrates an example of the wideband FEC processor **350**. FEC cabinet **410** stores the unquantized DFT of the residual signal (previously supplied by domain transformer **330**) along with quantization indices of the source model.

[0066] Wideband FEC processor **350** reads quantization indices of the source model previously stored in FEC cabinet **410** (e.g., quantization indices of pitch gain and pitch lag, quantization indices of LP-shape), and entropy codes them with an instance of entropy coder **340**, which has been described previously.

[0067] Wideband FEC processor **350** also reads the unquantized DFT of the residual signal from the FEC cabinet **410**, and multiplies the DFT with the FEC Scale in multiplier **415**, and subsequently quantizes the result in quantizer **335**.

[0068] Quantizing a signal introduces quantization errors. The errors depend on the quantization step size. More steps represent more possible values (assuming that the quantizer has a support which is sufficiently large with respect to the range of the signal), thus the actual value of the input signal is likely to be closer to one of the available values, thus resulting in a smaller quantization error. Thus, adjusting the quantization step size effects the quantization error. It is also possible to control the quantization error by maintaining a constant step size, and instead scaling the signal to be quantized. This is accomplished by the multiplication by the FEC scale. If the amplitude of the incoming signal is scaled with a large gain, more quantization steps are available to map to the actual values of the incoming signal (assuming that the quantizer has a support which is sufficiently large), again reducing the relative quantization error. If the incoming signal is scaled with a small gain, the result is higher quantization error, but also a lower encoding rate.

[0069] A higher encoding rate corresponds to finer quantization of the residual signal, while a lower encoding rate corresponds to coarser quantization. Therefore, the encoding rate is controlled by the value of the FEC scale, which is applied to the DFT of the residual

signal. A higher value result is a finer quantization (relatively smaller step-size) and higher encoding rate. The FEC scale is set to a value less than one, to decrease the coding rate.

[0070] Choosing the FEC scale is a compromise between concealment quality and the capacity which is consumed by redundant payloads. It is desirable to pick the FEC scale to provide good concealment quality in a situation that is likely to occur. Thus, it is possible to select the FEC scale as a function of the anticipated packet loss percentage, as long as the encoder and the decoder in two sides of the call agree on value of the packet loss. This might be communicated through in-band or out-of-band signaling. On the other hand, the FEC scale can also be pre-selected and hard-coded into the coder and decoder. Experiments by the inventors have shown that an FEC scale of 0.4 to 0.5 produces a good compromise between encoding rate and perceived audio quality where the packet loss rate is between 10% and 15%. Setting the FEC scale to 0.5 produces a redundant payload that is about half the size of the main payload.

[0071] It can be appreciated from the above, that given an input frame, the signal model and the residual signals are independent of the encoding rate. Hence, for encoding a frame at multiple rates, all complex operations of source modeling (e.g., LPC analysis, pitch estimation) and analysis filtering (analysis filter is inverse of production filter) are done only once. Then the residual signal (in proper domain, e.g., frequency domain) is stored together with the source model and re-used to obtain payloads at different encoding rates by simply multiplying the residual signal with a scale factor prior to quantization to obtain a payload bitstream at a desired encoding rate.

[0072] **Fig. 5** illustrates an example of an embodiment of the super-wideband encoder **103**. The super-wideband encoder **103** operates on the frequency band of 8-16 kHz. Optionally, this band may be downsampled to 8-12 kHz by down-sampler **501**. The super-wideband encoder **103** can have two modes. In the first mode only 8-12 kHz signal is encoded, while in the second mode the entire 8-16 kHz is encoded. The modes are chosen according to the available bit-rate. In both modes input and output are sampled at 32 kHz but only 0-12 kHz is encoded when the first mode is used. Thus, in the first mode output has bandwidth of 12 kHz sampled at 32 kHz.

[0073] The input signal, whether 8-16 kHz or 8-12 kHz, is processed in source model analysis **510**. As noted regarding source model analysis **310**, the source model analysis **510** produces a source model of the super-wideband signal, which is supplied to an instance of

quantizer **534**. Quantizer **534** is similar to quantizer **334**, as they both encode the source model using the same quantization technique, but with some differences. A quantizer is designed according to the statistics of the input signal. Source models of wideband and super-wideband are statistically different, thus the quantizers are specifically adapted for the respective statistics of the input signals they quantize.

[0074] The output of quantizer **534** is the quantized value of the source model, which is supplied as filter parameters of analysis filter **516**. Analysis filter **516** filters the incoming digital signal input based on the quantized values of the source model and outputs a residual signal of the incoming digital signal. Analysis filter **516** is similar to analysis filter **316**, but uses lower order LPC and includes more subframes as compared to the analysis filter **316** used for the 4-8kHz band.

[0075] An instance of entropy coder **340** entropy codes the quantized source model, outputting an entropy coded bitstream of the source model, which is supplied to super-wideband FEC processor **550**, and stored in FEC cabinet **610**. Output of the entropy coder **340** together with the state of the entropy coder is passed to FEC processor **550**. The entropy coded bitstream is also available for creating the payload for packets to be transmitted.

[0076] The residual signal output from analysis filter **516** is transformed in an instance of domain transformer **330**, and the transformed residual signal is supplied to super-wideband FEC processor **550** (e.g., as unquantized DFT coefficients) and stored in FEC cabinet **610**. The DFT of the residual signal is then quantized by an instance of quantizer **335**, and entropy coded by an instance of entropy coder **340**. The output of this instance of entropy coder **340** is an entropy-coded bitstream of the quantized residual signal, and the bitstream is available for creating payload for packets to be transmitted.

[0077] **Fig. 6** illustrates an example of an embodiment of super-wideband FEC processor **550**. FEC cabinet **610** stores entropy coded bitstream of the source model and unquantized DFT coefficients of the residual signal. For the super-wideband portion (8-16 kHz), the output of entropy coder **340** is stored in the FEC cabinet **610** together with state of the entropy coder. The output of the entropy coder is the actual bit-stream which constitutes a payload. Therefore, in the super-wideband FEC processor **550** this bitstream will be directly used to constitute the first segment of the super-wideband bit-stream. The state of the entropy

coder stored in FEC cabinet **610** is used to initialize the entropy coder **340** of FEC processor **550** and then the process continues as depicted.

[0078] The super-wideband FEC processor **550** retrieves the entropy coded bit stream of the source model and outputs it along with a redundant bitstream representing the residual signal. To generate the redundant bitstream, the unquantized DFT coefficients of the residual signal is retrieved from FEC cabinet **610** and multiplied by the FEC scale, which has been previously described. The FEC scale used in the super-wideband FEC processor **550** may be set to the same value as the FEC scale used in the wideband FEC processor **350**, or may be set to any other appropriate value. For example, the wideband FEC scale may be set to 0.4 and the super-wideband FEC scale can be set to 0.5.

[0079] The scaled residual signal is quantized by an instance of quantizer **335**, and then entropy coded by an instance of entropy coder **340**. It can be appreciated that the scaling and subsequent quantization effectively controls the coding rate of the redundant bit-stream by simply varying the FEC scale.

[0080] Entropy coder **340** can be implemented as a range encoder with a state, which is modified upon encoding a given index. As the main payload and redundant payload have the same source model (e.g., LP parameters), the source model is encoded only once, and stored in FEC cabinet **610**. When performing FEC processing, the stored bit-stream constitutes the first segment of the redundant payload, and the entropy coder **340** of the FEC processor **550** is initialized to the state was stored from the entropy coder **340** of the super-wideband encoder **103**.

[0081] At this stage, sufficient data is available to assemble packets containing information of the processed audio segments, both a main payload, and a redundant payload at a lower coding rate. The data can be thought of as encoded at some particular coding rate, which corresponds to a particular bandwidth. Thus, packets can be assembled with both a primary payload, and a redundant payload, in a forward error correction scheme. It is advantageous to encode the data for the redundant payload at a lower coding rate, thus using less bandwidth. The process of re-encoding at a lower rate would normally be computationally expensive, as the process would involve all of the processing described thus far. However, in the present invention the FEC processors **350** and **550** encode audio data at a lower rate which is a fraction of the primary coding rate, yet do not repeat the whole encoding process (including source modeling and analysis filtering).

[0082] Referring back to **Fig. 2**, both the wideband encoder **102** and super-wideband encoder **103** output bitstreams, which are concatenated and transmitted over an IP channel. The bitstreams are assembled into packets, which can include redundant payload for the FEC scheme. The redundant payload is created from the bitstreams provided by FEC processors **350** and **550**.

[0083] Packets for the FEC scheme include encoded data at two different coding rates. The primary rate corresponds to FEC scale equal to 1, while the encoding rate of the redundant data is some fraction (less than 1) of the primary coding rate, based on the value of the FEC scale set for the wideband FEC processor and the FEC scale set for the super-wideband FEC processor. The FEC scale is known both by encoders **102**, **103** and decoders **106**, **107**. Packets are assembled so that each packet contains encoded data corresponding to a time segment of audio encoded at the primary coding rate (primary payload) and also encoded data corresponding to an earlier time segment of audio data encoded at the redundant rate (redundant payload). This allows FEC using redundant payloads, while reducing required network bandwidth, and keeping the processing overhead for generating redundant payloads to a minimum. Thus, the coding rate is changed by changing the coarseness of quantization, which is a computationally inexpensive operation.

[0084] The preceding description illustrates an example of an embodiment of the redundant coding unit for an audio codec according to an embodiment of the invention. The following description along with **Fig. 7** describes a high level processing flow of generating redundant bitstream for encoded audio data.

[0085] In step **S710**, the incoming audio signal is analyzed to obtain the source model and the residual signal. The details vary for the wideband portion and the super-wideband portion, and further for the low band and high band of the wideband portion. The source model and the residual signal can be output as a bitstream for creating of the primary payload of data packets.

[0086] In step **S720**, the representation of the source model and the residual signal are stored in respective FEC cabinets for wideband and super-wideband portions. The encoding of the source model is independent of the coding rate, while the residual signal is scaled by the FEC scale to control the coding rate.

[0087] In step **S730**, the previously stored representations of the source model and the residual signal are retrieved, and the residual signal is scaled by the FEC scale. The scaled

residual signal is then quantized in step **S740**, which effectively controls the coding rate. The quantized residual signal is also entropy coded, and in step **S750** the bitstream for forming the redundant payload of data packets is output.

[0088] As illustrated in **Fig. 2**, bitstreams generated by the wideband encoder **102** and super-wideband encoder **103** are concatenated, and data packets are formed. The data packets may contain a primary payload encoded at a primary rate, and a redundant payload encoded at a lower rate set by the FEC scale. A packet may thus contain primary payload for time segment n and a redundant payload for time segment $n-1$. It is also possible to have additional redundant payloads for other time segments coded at various coding rates, but the example below is directed to one primary payload and one redundant payload. The data packets are transmitted to the receiving side, where the bitstreams of encoded audio data are separated from the data packets and are separated for the wideband and the super-wideband portion.

[0089] The wideband portion of the bitstream is decoded by wideband decoder **106**, which is illustrated in **Fig. 8**. The super-wideband portion is decoded by super-wideband decoder **107**, illustrated in **Fig. 9**.

[0090] Entropy decoder **822** of the wideband decoder **106** receives the bitstream, decodes it, and outputs quantization indices of the source model and the residual signal. The source model is supplied to source model decoder **810**, which includes de-quantizer **814**.

[0091] The output of de-quantizer **814** represents the source model for each of the low band (0-4kHz) and the high band (4-8kHz), and provides filter coefficients for the synthesis filters **835** and **836**.

[0092] The quantization indices of the residual signal output from entropy decoder **822** are provided to de-quantizer **815** of the spectrum decoder **820**, which de-quantizes the residual signal (resulting in a DFT coefficients of the residual signal).

[0093] In the case of redundant payload that has been scaled by the FEC scale, the DFT of the residual signal is divided by the same FEC scale factor in divider **818** as used when creating the redundant payload. The FEC scale is known by both the encoder and decoder. In the case of primary payload that has not been scaled, the division is skipped (effectively dividing by one).

[0094] The scaled DFT of the residual signal is transformed back into the time domain by inverse domain transformer **830**, and supplied to synthesis filters **835** and **836**. The inverse

domain transformer **830** may be implemented as an IFFT (Inverse Fast Fourier Transform), and transforms the residual signal into the time domain.

[0095] The processing for the low band and the high band can vary in the synthesis filters **835** and **836**. In the low band (0-4 kHz), synthesis filter **835** may perform the job of synthesis filtering. Like its analysis counterpart (analysis filter **315**), the synthesis filter **835** can be implemented in two steps. The first step is a pole-zero filter derived from pitch gain and lag, reconstructing long-term dependencies. The second step is an all-pole filter derived from LPC parameters. For the high band (4-8 kHz), the synthesis filter **836** can be derived from LPC and is an all-pole filter. All-pole synthesis filters may be implemented as lattice filters. The synthesis filter **835** is linear quasi-time-invariant filter, where filter coefficients are updated with the rate that LP parameters are updated.

[0096] Analysis filter **315** and synthesis filter **835** are inverse of each other (within the accuracy of the implementation), such that if there were no quantization of the DFT coefficients, the input to analysis filter **315** (shown in **Fig. 3**) would be the same as the output of synthesis filter **835** (shown in **Fig. 8**). The same holds for analysis filter **316** and synthesis filter **836**.

[0097] The output of synthesis filters **835** is the audio signal for each of low band and high band, which is then combined in filter-bank **840** into the wideband signal (0-8kHz). Filter-bank **840** is the inverse of filter-bank **302**. Thus, if there were no quantization in the path from filter-bank **302** to filter-bank **840**, then the reconstructed signal would be the same as the input (within the accuracy of the implementation).

[0098] The super-wideband decoder **107** operates in a similar fashion, but processes the entire signal (8-16 kHz, or 8-12kHz) together, as illustrated in **Fig. 9**. The source model decoder **910** includes de-quantizer **914**. The de-quantizer **914** differs from the de-quantizer **814** as their counterpart quantizers (**334** and **534**, respectively) are different. In general, a pair of quantizer and de-quantizer should match each other for the best reconstruction of the source. For instance, a pair of uniform scalar quantizer and dequantizer should have the same steps size. The output of the synthesis filter **936** of the super-wideband decoder **107** is optionally upsampled by upsampler **920** (when 8-12 kHz as the bandwidth) is used. Analysis filter **516** and synthesis filter **936** are inverse of each other (within the accuracy of the implementation), such that if there was no quantization of the DFT coefficients, the input to

analysis filter **516** (shown in **Fig. 5**) would be the same as the output of synthesis filter **936** (shown in **Fig. 9**).

[0099] **Fig. 10** illustrates an example of a high level processing flow of decoding a redundant payload for frame *n* corresponding to a time segment of audio data.

[0100] In step **S1010**, a packet with a primary payload and a redundant payload is received.

[0101] In step **S1020**, the redundant payload is extracted from the packet. The encoded residual signal is separated from the encoded source model (e.g., LP shape parameters).

[0102] In step **S1030**, the residual signal is entropy decoded and dequantized.

[0103] In step **S1040**, the residual signal is divided by the FEC scale.

[0104] In step **S1050**, the residual signal is transformed into the time domain by inverse domain transformer **830**.

[0105] In step **S1060**, the residual signal is filtered by the synthesis filter, based on the decoded source model, to recreate a digital representation of the transmitted signal.

[0106] In step **S1070**, the digital representation is optionally upsampled.

[0107] In step **S1080**, the signal is converted to an analog signal.

[0108] **FIG. 11** is a block diagram illustrating an example of a computing device **1100** that is arranged for performing redundant coding and decoding in accordance with the present disclosure. In a very basic configuration **1101**, computing device **1100** typically includes one or more processors **1110** and system memory **1120**. A memory bus **1130** can be used for communicating between the processor **1110** and the system memory **1120**.

[0109] Depending on the desired configuration, processor **1110** can be of any type including but not limited to a microprocessor (μ P), a microcontroller (μ C), a digital signal processor (DSP), or any combination thereof. Processor **1110** can include one more levels of caching, such as a level one cache **1111** and a level two cache **1112**, a processor core **1113**, and registers **1114**. The processor core **1113** can include an arithmetic logic unit (ALU), a floating-point unit (FPU), a digital signal processing core (DSP Core), or any combination thereof. A memory controller **1115** can also be used with the processor **1110**, or in some implementations the memory controller **1115** can be an internal part of the processor **1110**.

[0110] Depending on the desired configuration, the system memory **1120** can be of any type including but not limited to volatile memory (such as RAM), non-volatile memory (such as ROM, flash memory, etc.) or any combination thereof. System memory **1120** typically

includes an operating system **1121**, one or more applications **1122**, and program data **1124**. Application **1122** includes a coding and decoding algorithm with FEC support **1123** that is arranged to perform the coding and decoding as described in this disclosure. Program Data **1124** includes service data **1125** that is useful for performing coding and decoding of audio signals, as will be further described below. In some embodiments, application **1122** can be arranged to operate with program data **1124** on an operating system **1121**. This described basic configuration is illustrated in **FIG. 11** by those components within dashed line **1101**.

[0111] Computing device **1100** can have additional features or functionality, and additional interfaces to facilitate communications between the basic configuration **1101** and any required devices and interfaces. For example, a bus/interface controller **1140** can be used to facilitate communications between the basic configuration **1101** and one or more data storage devices **1150** via a storage interface bus **1141**. The data storage devices **1150** can be removable storage devices **1151**, non-removable storage devices **1152**, or a combination thereof. Examples of removable storage and non-removable storage devices include magnetic disk devices such as flexible disk drives and hard-disk drives (HDD), optical disk drives such as compact disk (CD) drives or digital versatile disk (DVD) drives, solid state drives (SSD), and tape drives to name a few. Example computer storage media can include volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information, such as computer readable instructions, data structures, program modules; or other data.

[0112] System memory **1120**, removable storage **1151** and non-removable storage **1152** are all examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by computing device **1100**. Any such computer storage media can be part of device **1100**.

[0113] Computing device **1100** can also include an interface bus **1142** for facilitating communication from various interface devices (e.g., output interfaces, peripheral interfaces, and communication interfaces) to the basic configuration **1101** via the bus/interface controller **1140**. Example output devices **1160** include a graphics processing unit **1161** and an audio processing unit **1162**, which can be configured to communicate to various external devices

such as a display or speakers via one or more A/V ports **1163**. Example peripheral interfaces **1170** include a serial interface controller **1171** or a parallel interface controller **1172**, which can be configured to communicate with external devices such as input devices (e.g., keyboard, mouse, pen, voice input device, touch input device, etc.) or other peripheral devices (e.g., printer, scanner, etc.) via one or more I/O ports **1173**. An example communication device **1180** includes a network controller **1181**, which can be arranged to facilitate communications with one or more other computing devices **1190** over a network communication via one or more communication ports **1182**. The communication connection is one example of a communication media. Communication media may typically be embodied by computer readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave or other transport mechanism, and includes any information delivery media. A “modulated data signal” can be a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media can include wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, radio frequency (RF), infrared (IR) and other wireless media. The term computer readable media as used herein can include both storage media and communication media.

[0114] Computing device **1100** can be implemented as a portion of a small-form factor portable (or mobile) electronic device such as a cell phone, a personal data assistant (PDA), a personal media player device, a wireless web-watch device, a personal headset device, an application specific device, or a hybrid device that include any of the above functions. Computing device **1100** can also be implemented as a personal computer including both laptop computer and non-laptop computer configurations.

[0115] There is little distinction left between hardware and software implementations of aspects of systems; the use of hardware or software is generally (but not always, in that in certain contexts the choice between hardware and software can become significant) a design choice representing cost vs. efficiency tradeoffs. There are various vehicles by which processes and/or systems and/or other technologies described herein can be effected (e.g., hardware, software, and/or firmware), and that the preferred vehicle will vary with the context in which the processes and/or systems and/or other technologies are deployed. For example, if an implementer determines that speed and accuracy are paramount, the implementer may opt for a mainly hardware and/or firmware vehicle; if flexibility is

paramount, the implementer may opt for a mainly software implementation; or, yet again alternatively, the implementer may opt for some combination of hardware, software, and/or firmware.

[0116] The foregoing detailed description has set forth various embodiments of the devices and/or processes via the use of block diagrams, flowcharts, and/or examples. Insofar as such block diagrams, flowcharts, and/or examples contain one or more functions and/or operations, it will be understood by those within the art that each function and/or operation within such block diagrams, flowcharts, or examples can be implemented, individually and/or collectively, by a wide range of hardware, software, firmware, or virtually any combination thereof. In one embodiment, several portions of the subject matter described herein may be implemented via Application Specific Integrated Circuits (ASICs), Field Programmable Gate Arrays (FPGAs), digital signal processors (DSPs), or other integrated formats. However, those skilled in the art will recognize that some aspects of the embodiments disclosed herein, in whole or in part, can be equivalently implemented in integrated circuits, as one or more computer programs running on one or more computers (e.g., as one or more programs running on one or more computer systems), as one or more programs running on one or more processors (e.g., as one or more programs running on one or more microprocessors), as firmware, or as virtually any combination thereof, and that designing the circuitry and/or writing the code for the software and or firmware would be well within the skill of one of skill in the art in light of this disclosure. In addition, those skilled in the art will appreciate that the mechanisms of the subject matter described herein are capable of being distributed as a program product in a variety of forms, and that an illustrative embodiment of the subject matter described herein applies regardless of the particular type of signal bearing medium used to actually carry out the distribution. Examples of a signal bearing medium include, but are not limited to, the following: a recordable type medium such as a floppy disk, a hard disk drive, a Compact Disc (CD), a Digital Video Disk (DVD), a digital tape, a computer memory, etc.; and a transmission type medium such as a digital and/or an analog communication medium (e.g., a fiber optic cable, a waveguide, a wired communications link, a wireless communication link, etc.).

[0117] Those skilled in the art will recognize that it is common within the art to describe devices and/or processes in the fashion set forth herein, and thereafter use engineering practices to integrate such described devices and/or processes into data processing systems.

That is, at least a portion of the devices and/or processes described herein can be integrated into a data processing system via a reasonable amount of experimentation. Those having skill in the art will recognize that a typical data processing system generally includes one or more of a system unit housing, a video display device, a memory such as volatile and non-volatile memory, processors such as microprocessors and digital signal processors, computational entities such as operating systems, drivers, graphical user interfaces, and applications programs, one or more interaction devices, such as a touch pad or screen, and/or control systems including feedback loops and control motors (e.g., feedback for sensing position and/or velocity; control motors for moving and/or adjusting components and/or quantities). A typical data processing system may be implemented utilizing any suitable commercially available components, such as those typically found in data computing/communication and/or network computing/communication systems.

[0118] With respect to the use of substantially any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity.

[0119] While various aspects and embodiments have been disclosed herein, other aspects and embodiments will be apparent to those skilled in the art. The various aspects and embodiments disclosed herein are for purposes of illustration and are not intended to be limiting, with the true scope and spirit being indicated by the following claims.

CLAIMS

1. A method of generating multiple payloads encoded at different coding rates for inclusion in a data packet transmitted across a packet data network, the data packet containing an encoded audio source signal, the method comprising:

separating the audio source signal into a first frequency band signal and a second frequency band signal, the first frequency band being lower than the second frequency band;

extracting from the first frequency band signal a source model representing linear dependencies of the first frequency band signal;

generating a residual signal by filtering the first frequency band signal with a filter having filter coefficients derived from the source model to remove short-term and long-term linear dependencies from the first frequency band signal;

transforming the residual signal into a transform domain;

scaling the transformed residual signal with a first scale factor;

quantizing the scaled transformed residual signal;

quantizing the source model to create quantization indices of the source model;

entropy encoding the quantization indices of the scaled transformed residual signal and the quantization indices of the source model to generate a redundant bitstream; and

constructing the payloads from the redundant bitstream.

2. The method according to claim 1, further comprising:

extracting from the second frequency band signal a second source model representing linear dependencies of the second frequency band signal;

generating a second residual signal by filtering the second frequency band signal with filter coefficients derived from the second source model to remove linear dependencies from the second frequency band signal;

transforming the second residual signal into the transform domain;

scaling the transformed second residual signal with a second scale factor;

quantizing the scaled transformed second residual signal to create quantization indices of the transformed residual signal;

quantizing the second source model to create quantization indices of the source model;

entropy encoding the quantization indices of the scaled transformed second residual signal and the quantization indices of the second source model to generate a second redundant bitstream; and

constructing the payloads from the second redundant bitstream.

3. The method according to claim 2, further comprising
storing in a memory the transformed residual signal and the quantization indices of the source model prior to the scaling of the transformed residual signal; and

extracting from the memory the stored transformed residual signal and the quantization indices of the source model when the redundant payload is to be encoded at a lower coding rate than a coding rate of a primary payload.

4. The method according to claim 1, wherein the step of separating the audio source signal into the first frequency band signal and the second frequency band signal includes:

dividing the first frequency band into a first sub-band and a second sub-band, the first sub-band being lower than the second sub-band.

5. The method according to claim 2, further comprising:

generating a data frame including

an encoded segment of the audio source signal corresponding to a first time period, and

an encoded segment of the audio source signal corresponding to a second time period different than the first time period, the second time period preceding the first time period.

6. The method according to claim 5, wherein

the segment of the audio source signal corresponding to the first time period is encoded at a higher coding rate than the coding rate of the segment of audio source data corresponding to the second time period.

7. The method according to claim 6, wherein

the first scale factor and the second scale factor are set independently of each other, based on the coding rate of the segment of audio source data corresponding to the second time period.

8. The method according to claim 7, wherein

the first scale factor and the second scale factor are each independently set to a value between 0.4 and 0.5, inclusive, to reach a suitable overall-quality vs concealment-quality trade off, at packet loss rates in the range of 10% to 15%, inclusive.

9. The method according to claim 2, further comprising:

downsampling the second frequency band signal prior to the extracting of the second source model.

10. An encoding apparatus for encoding a source audio signal at different coding rates to generate multiple payloads included in data packets transmitted across a packet data network, the apparatus comprising:

a filter-bank configured to separate the audio source signal into a first frequency band signal and a second frequency band signal, the first frequency band being lower than the second frequency band;

a source model analysis unit configured to generate a source model representing linear dependencies of the first frequency band signal;

an analysis filter having its filter coefficients derived from the source model and configured to filter the first frequency band signal to generate a residual signal;

a domain transformer transforming the residual signal into a transform domain;

a multiplier multiplying the transformed residual signal with a first scale factor;

a quantizer quantizing the scaled transformed residual signal, and quantizing the source model to create associated quantization indices of the source model and the scaled transformed residual signal; and

an entropy coder encoding the quantization indices of the scaled transformed residual signal and the quantization indices of the source model to generate a redundant bitstream.

11. The encoding apparatus according to claim 10, further comprising:

a second source model analysis unit configured to extract from the second frequency band signal a second source model representing linear dependencies of the second frequency band signal;

a second analysis filter having filter coefficients derived from the second source model and configured to filter the second frequency band signal to generate a second residual signal;

a second domain transformer transforming the second residual signal into the transform domain;

a second multiplier multiplying the transformed second residual signal with a second scale factor;

a second quantizer quantizing the scaled transformed second residual signal, and quantizing the second source model to create quantization indices of the second source model and quantization indices of the scaled transformed second residual signal; and

a second entropy coder encoding the quantization indices of the scaled transformed second residual signal and the quantization indices of the second source model to generate a second redundant bitstream.

12. The encoding apparatus according to claim 11, further comprising:

a storage unit storing the transformed residual signal and the quantization indices of the source model prior to the multiplication by the multiplier, wherein

the stored transformed residual signal and the quantization indices of the source model are extracted from the storage unit when the redundant payload is to be encoded at a lower coding rate than a coding rate of a primary payload.

13. The encoding apparatus according to claim 10, wherein

the filter bank is further configured to divide the first frequency band into a first sub-band and a second sub-band, the first sub-band being lower than the second sub-band.

14. The encoding apparatus according to claim 10, further comprising:

a concatenation unit configured to generate a data frame including an encoded segment of the audio source signal corresponding to a first time period and an encoded

segment of the audio source signal corresponding to a second time period different than the first time period, the second time period preceding the first time period.

15. The encoding apparatus according to claim 14, wherein the segment of the audio source signal corresponding to the first time period is encoded at a higher coding rate than the coding rate of the segment of audio source data corresponding to the second time period.

16. The encoding apparatus according to claim 15, wherein the first scale factor and the second scale factor are set independently of each other, based on the coding rate of the segment of audio source data corresponding to the second time period.

17. The encoding apparatus according to claim 16, wherein the first scale factor and the second scale factor are each independently set to a value between 0.4 and 0.5, inclusive, to reach a suitable overall-quality vs concealment-quality trade off, at packet loss rates in the range of 10% to 15%, inclusive.

18. The encoding apparatus according to claim 11, further comprising: a downsampler configured to downsample the second frequency band signal prior to processing by the second source model analysis unit and the second analysis filter.

19. A computer readable tangible recording medium encoded with instructions, wherein the instructions when executed by a processor cause the processor to perform a method of generating multiple payloads encoded at different coding rates for inclusion in a data packet transmitted across a packet data network, the method comprising:
separating the audio source signal into a first frequency band signal and a second frequency band signal, the first frequency band being lower than the second frequency band;
extracting from the first frequency band signal a source model representing linear dependencies of the first frequency band signal;

generating a residual signal by filtering the first frequency band signal with a filter having filter coefficients derived from the source model to remove short-term and long-term linear dependencies from the first frequency band signal;
transforming the residual signal into a transform domain;
scaling the transformed residual signal with a first scale factor;
quantizing the scaled transformed residual signal;
quantizing the source model to create quantization indices of the source model;
entropy encoding the quantization indices of the scaled transformed residual signal and the quantization indices of the source model to generate a redundant bitstream; and
constructing the payloads from the redundant bitstream.

20. The computer readable tangible recording medium according to claim 19, wherein the method further comprises:

extracting from the second frequency band signal a second source model representing linear dependencies of the second frequency band signal;
generating a second residual signal by filtering the second frequency band signal with filter coefficients derived from the second source model to remove linear dependencies from the second frequency band signal;
transforming the second residual signal into the transform domain;
scaling the transformed second residual signal with a second scale factor;
quantizing the scaled transformed second residual signal to create quantization indices of the transformed residual signal;
quantizing the second source model to create quantization indices of the source model;
entropy encoding the quantization indices of the scaled transformed second residual signal and the quantization indices of the second source model to generate a second redundant bitstream; and
constructing the payloads from the second redundant bitstream.

AMENDED CLAIMS
received by the International Bureau on 17 December 2011 (17.12.2011)

Claims

1. A method of generating multiple payloads encoded at different coding rates for performing forward error correction (FEC) based on redundant data included in a data packet transmitted across a packet data network, the data packet containing an encoded audio source signal, the method comprising:

separating the audio source signal into a first frequency band signal and a second frequency band signal, the first frequency band being lower than the second frequency band;

extracting from the first frequency band signal a source model representing linear dependencies of the first frequency band signal;

generating a residual signal by filtering the first frequency band signal with a filter having filter coefficients derived from the source model to remove short-term and long-term linear dependencies from the first frequency band signal;

transforming the residual signal into a transform domain;

scaling the transformed residual signal with a first scale factor;

quantizing the scaled transformed residual signal;

quantizing the source model to create quantization indices of the source model;

entropy encoding the quantization indices of the scaled transformed residual signal and the quantization indices of the source model to generate a redundant bitstream; and
constructing FEC payloads from the redundant bitstream.

2. The method according to claim 1, further comprising:

extracting from the second frequency band signal a second source model representing linear dependencies of the second frequency band signal;

generating a second residual signal by filtering the second frequency band signal with filter coefficients derived from the second source model to remove linear dependencies from the second frequency band signal;

transforming the second residual signal into the transform domain;

scaling the transformed second residual signal with a second scale factor;

quantizing the scaled transformed second residual signal to create quantization indices of the transformed residual signal;

quantizing the second source model to create quantization indices of the source model;

entropy encoding the quantization indices of the scaled transformed second residual signal and the quantization indices of the second source model to generate a second redundant bitstream; and

constructing the FEC payloads from the second redundant bitstream.

3. The method according to claim 2, further comprising
storing in a memory the transformed residual signal and the quantization indices of the source model prior to the scaling of the transformed residual signal; and
extracting from the memory the stored transformed residual signal and the quantization indices of the source model when the redundant payload is to be encoded at a lower coding rate than a coding rate of a primary payload.

4. The method according to claim 1, wherein the step of separating the audio source signal into the first frequency band signal and the second frequency band signal includes:
dividing the first frequency band into a first sub-band and a second sub-band, the first sub-band being lower than the second sub-band.

5. The method according to claim 2, further comprising:
generating a data frame including
an encoded segment of the audio source signal corresponding to a first time period, and
an FEC payload including an encoded segment of the audio source signal corresponding to a second time period different than the first time period, the second time period preceding the first time period.

6. The method according to claim 5, wherein
the segment of the audio source signal corresponding to the first time period is encoded at a higher coding rate than the coding rate of the segment of audio source data corresponding to the second time period.

7. The method according to claim 6, wherein the first scale factor and the second scale factor are set independently of each other, based on the coding rate of the segment of audio source data corresponding to the second time period.

8. The method according to claim 7, wherein the first scale factor and the second scale factor are each independently set to a value between 0.4 and 0.5, inclusive, at packet loss rates in the range of 10% to 15%, inclusive.

9. The method according to claim 2, further comprising: downsampling the second frequency band signal prior to the extracting of the second source model.

10. An encoding apparatus for encoding a source audio signal at different coding rates to generate multiple payloads included in data packets transmitted with forward error correction (FEC) based on redundant data across a packet data network, the apparatus comprising:

a filter-bank configured to separate the audio source signal into a first frequency band signal and a second frequency band signal, the first frequency band being lower than the second frequency band;

a source model analysis unit configured to generate a source model representing linear dependencies of the first frequency band signal;

an analysis filter having its filter coefficients derived from the source model and configured to filter the first frequency band signal to generate a residual signal;

a domain transformer transforming the residual signal into a transform domain;

a multiplier multiplying the transformed residual signal with a first scale factor;

a quantizer quantizing the scaled transformed residual signal, and quantizing the source model to create associated quantization indices of the source model and the scaled transformed residual signal; and

an entropy coder encoding the quantization indices of the scaled transformed residual signal and the quantization indices of the source model to generate a redundant bitstream includable as a part of an FEC payload.

11. The encoding apparatus according to claim 10, further comprising:
a second source model analysis unit configured to extract from the second frequency band signal a second source model representing linear dependencies of the second frequency band signal;
a second analysis filter having filter coefficients derived from the second source model and configured to filter the second frequency band signal to generate a second residual signal;
a second domain transformer transforming the second residual signal into the transform domain;
a second multiplier multiplying the transformed second residual signal with a second scale factor;
a second quantizer quantizing the scaled transformed second residual signal, and quantizing the second source model to create quantization indices of the second source model and quantization indices of the scaled transformed second residual signal; and
a second entropy coder encoding the quantization indices of the scaled transformed second residual signal and the quantization indices of the second source model to generate a second redundant bitstream includable as another part of the FEC payload.

12. The encoding apparatus according to claim 11, further comprising:
a storage unit storing the transformed residual signal and the quantization indices of the source model prior to the multiplication by the multiplier, wherein
the stored transformed residual signal and the quantization indices of the source model are extracted from the storage unit when the redundant payload is to be encoded at a lower coding rate than a coding rate of a primary payload.

13. The encoding apparatus according to claim 10, wherein
the filter bank is further configured to divide the first frequency band into a first sub-band and a second sub-band, the first sub-band being lower than the second sub-band.

14. The encoding apparatus according to claim 10, further comprising:

a concatenation unit configured to generate a data frame including an encoded segment of the audio source signal corresponding to a first time period and an FEC payload including an encoded segment of the audio source signal corresponding to a second time period different than the first time period, the second time period preceding the first time period,

15. The encoding apparatus according to claim 14, wherein the segment of the audio source signal corresponding to the first time period is encoded at a higher coding rate than the coding rate of the segment of audio source data corresponding to the second time period.

16. The encoding apparatus according to claim 15, wherein the first scale factor and the second scale factor are set independently of each other, based on the coding rate of the segment of audio source data corresponding to the second time period.

17. The encoding apparatus according to claim 16, wherein the first scale factor and the second scale factor are each independently set to a value between 0.4 and 0.5, inclusive, at packet loss rates in the range of 10% to 15%, inclusive.

18. The encoding apparatus according to claim 11, further comprising:
a downsampler configured to downsample the second frequency band signal prior to processing by the second source model analysis unit and the second analysis filter.

19. A computer readable tangible recording medium encoded with instructions, wherein the instructions when executed by a processor cause the processor to perform a method of generating multiple payloads encoded at different coding rates for performing forward error correction (FEC) based on redundant data included in a data packet transmitted across a packet data network, the method comprising:

separating the audio source signal into a first frequency band signal and a second frequency band signal, the first frequency band being lower than the second frequency band;

extracting from the first frequency band signal a source model representing linear dependencies of the first frequency band signal;

generating a residual signal by filtering the first frequency band signal with a filter having filter coefficients derived from the source model to remove short-term and long-term linear dependencies from the first frequency band signal;

transforming the residual signal into a transform domain;

scaling the transformed residual signal with a first scale factor;

quantizing the scaled transformed residual signal;

quantizing the source model to create quantization indices of the source model;

entropy encoding the quantization indices of the scaled transformed residual signal and the quantization indices of the source model to generate a redundant bitstream; and

constructing FEC payloads from the redundant bitstream.

20. The computer readable tangible recording medium according to claim 19, wherein the method further comprises:

extracting from the second frequency band signal a second source model representing linear dependencies of the second frequency band signal;

generating a second residual signal by filtering the second frequency band signal with filter coefficients derived from the second source model to remove linear dependencies from the second frequency band signal;

transforming the second residual signal into the transform domain;

scaling the transformed second residual signal with a second scale factor;

quantizing the scaled transformed second residual signal to create quantization indices of the transformed residual signal;

quantizing the second source model to create quantization indices of the source model;

entropy encoding the quantization indices of the scaled transformed second residual signal and the quantization indices of the second source model to generate a second redundant bitstream; and

constructing the FEC payloads from the second redundant bitstream.

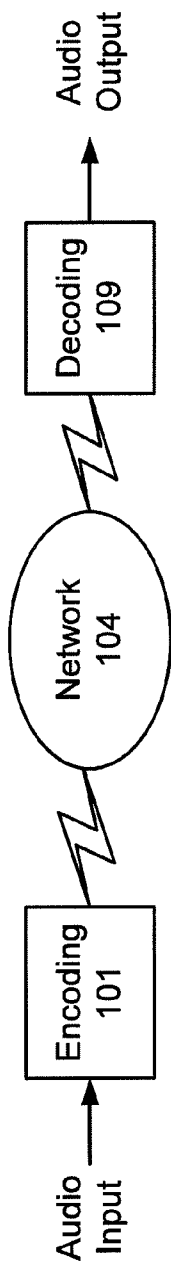


FIG. 1

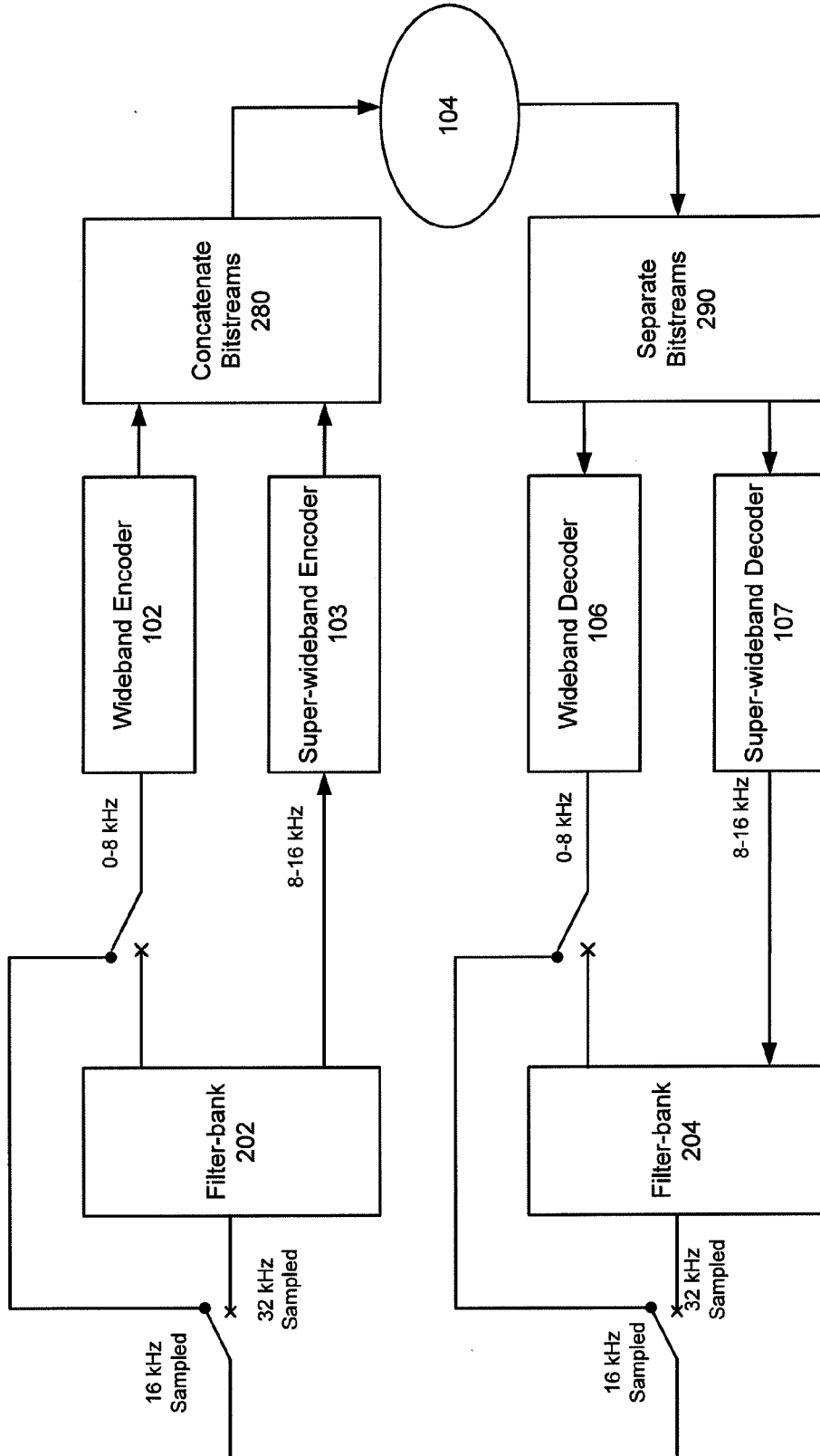


FIG. 2

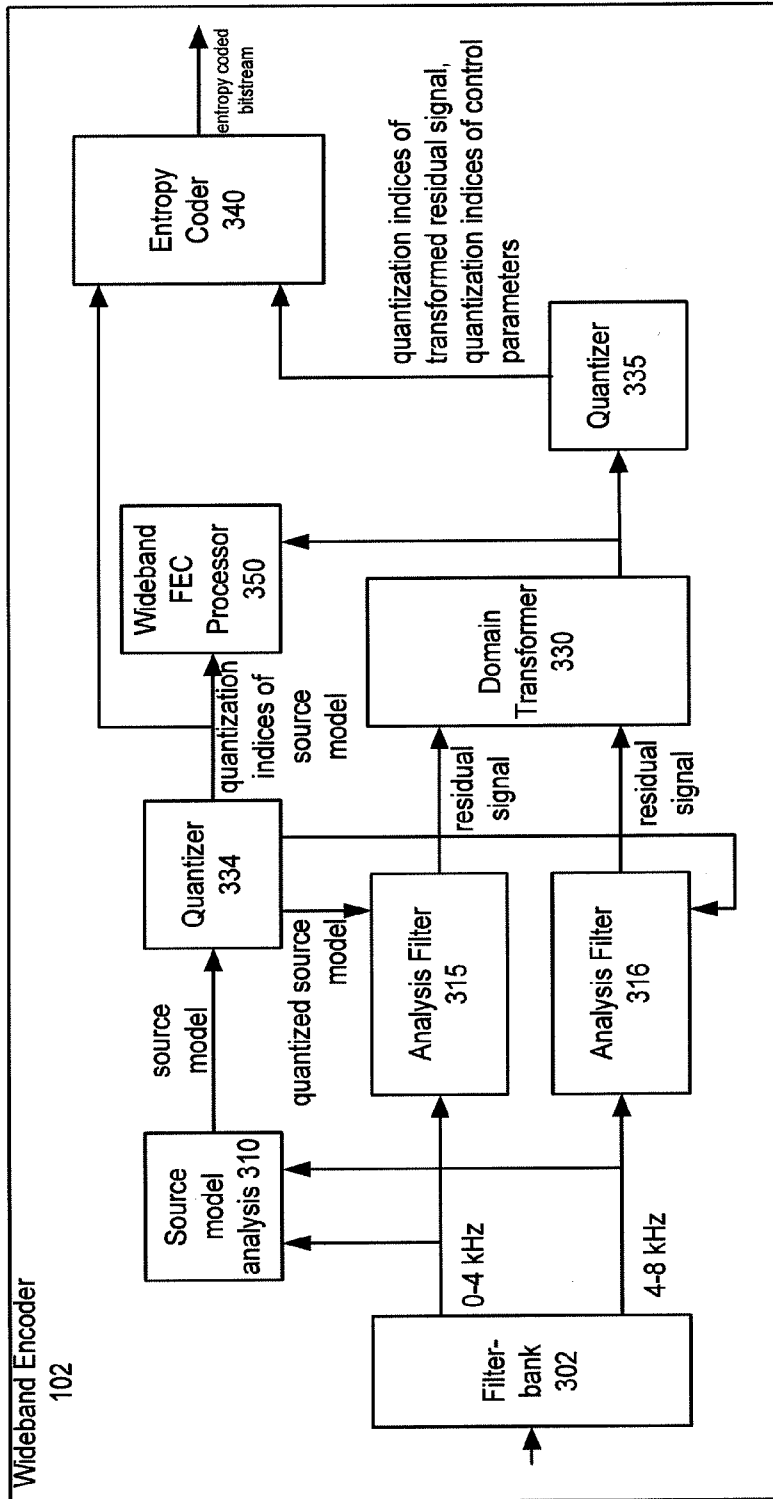


FIG. 3

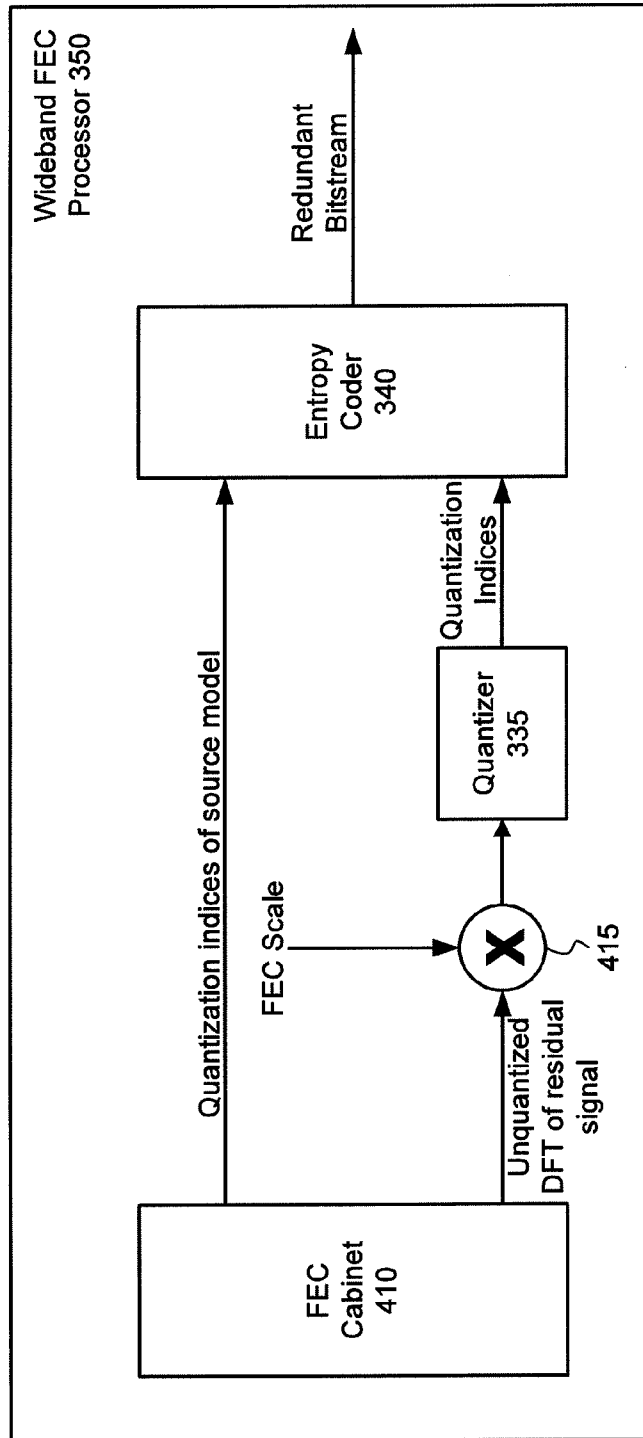


FIG. 4

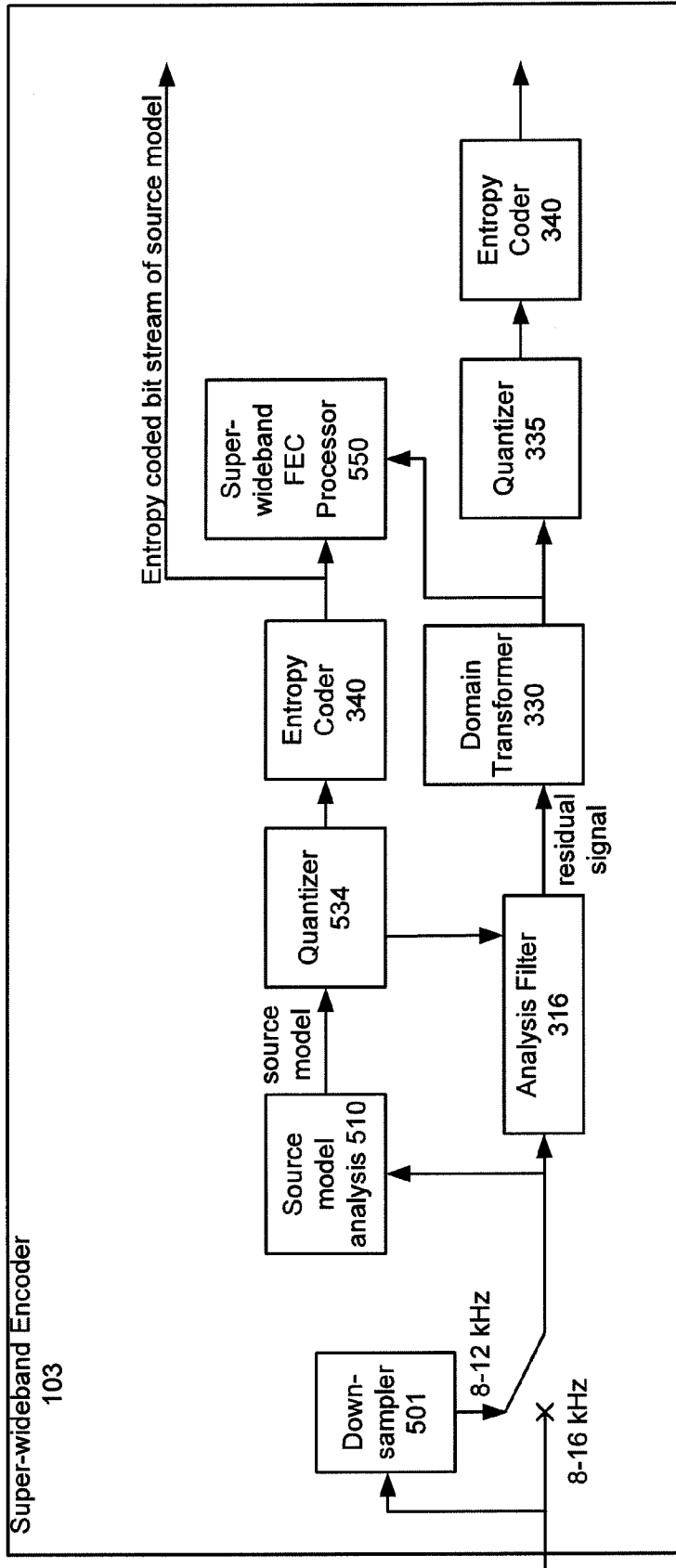


FIG. 5

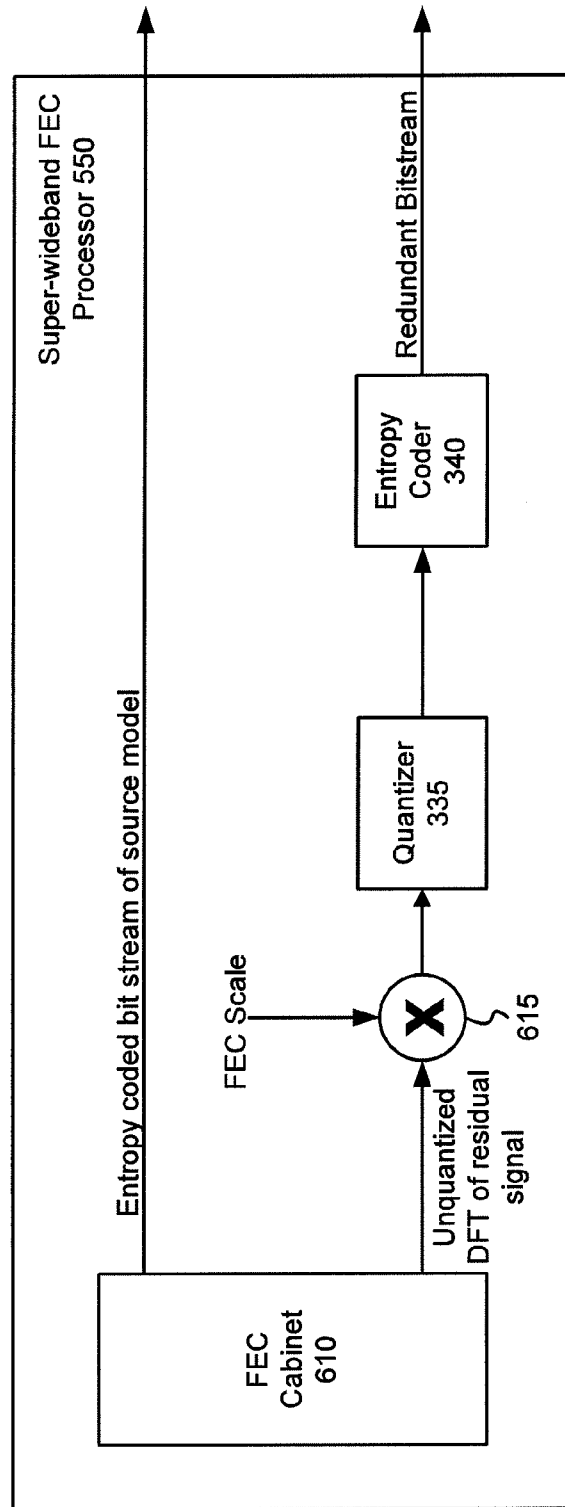


FIG. 6

7/11

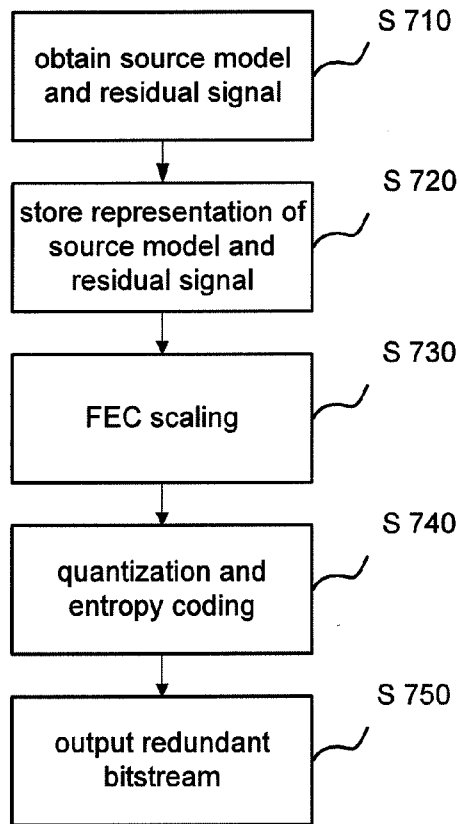


FIG. 7

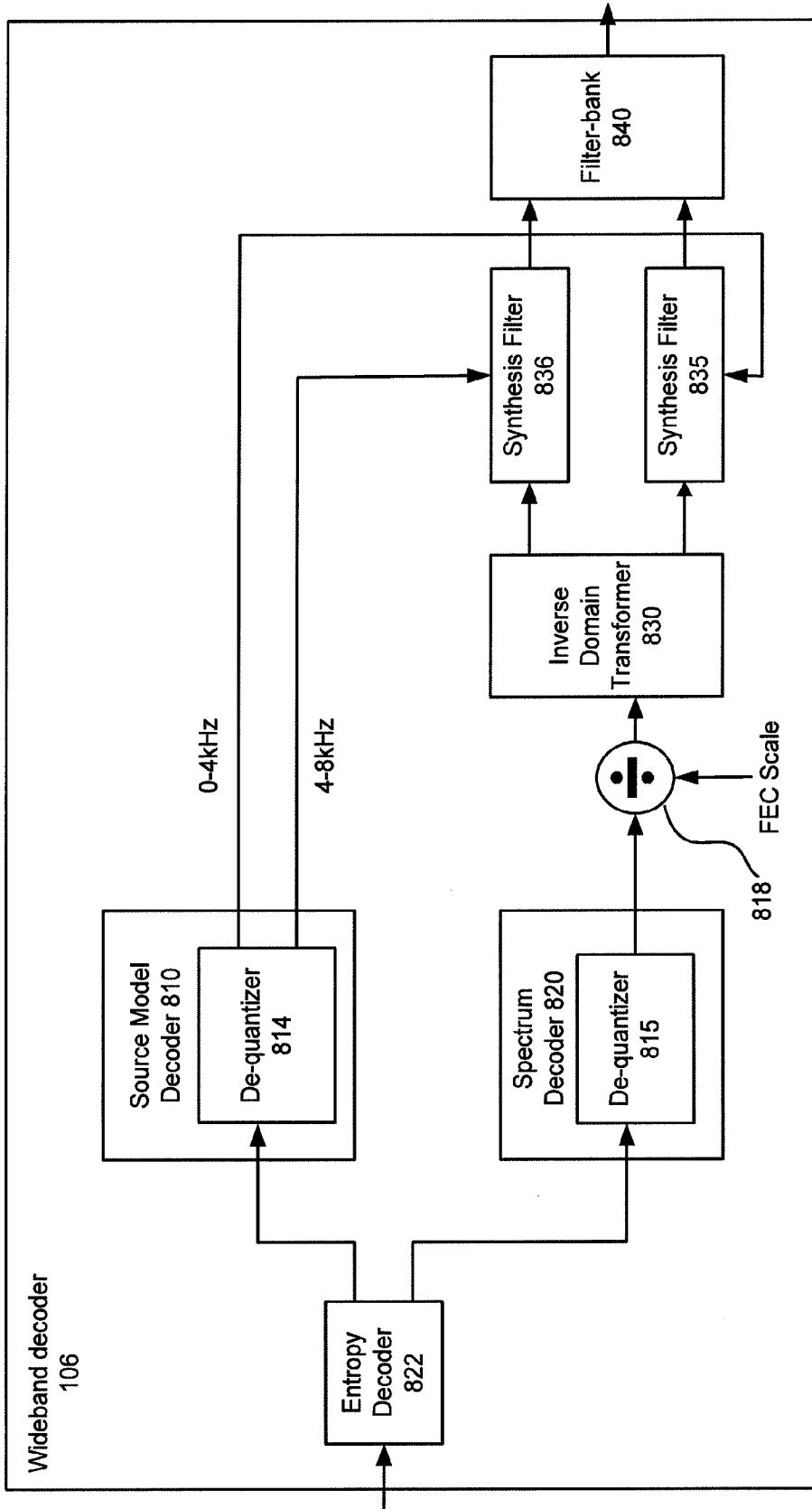


FIG. 8

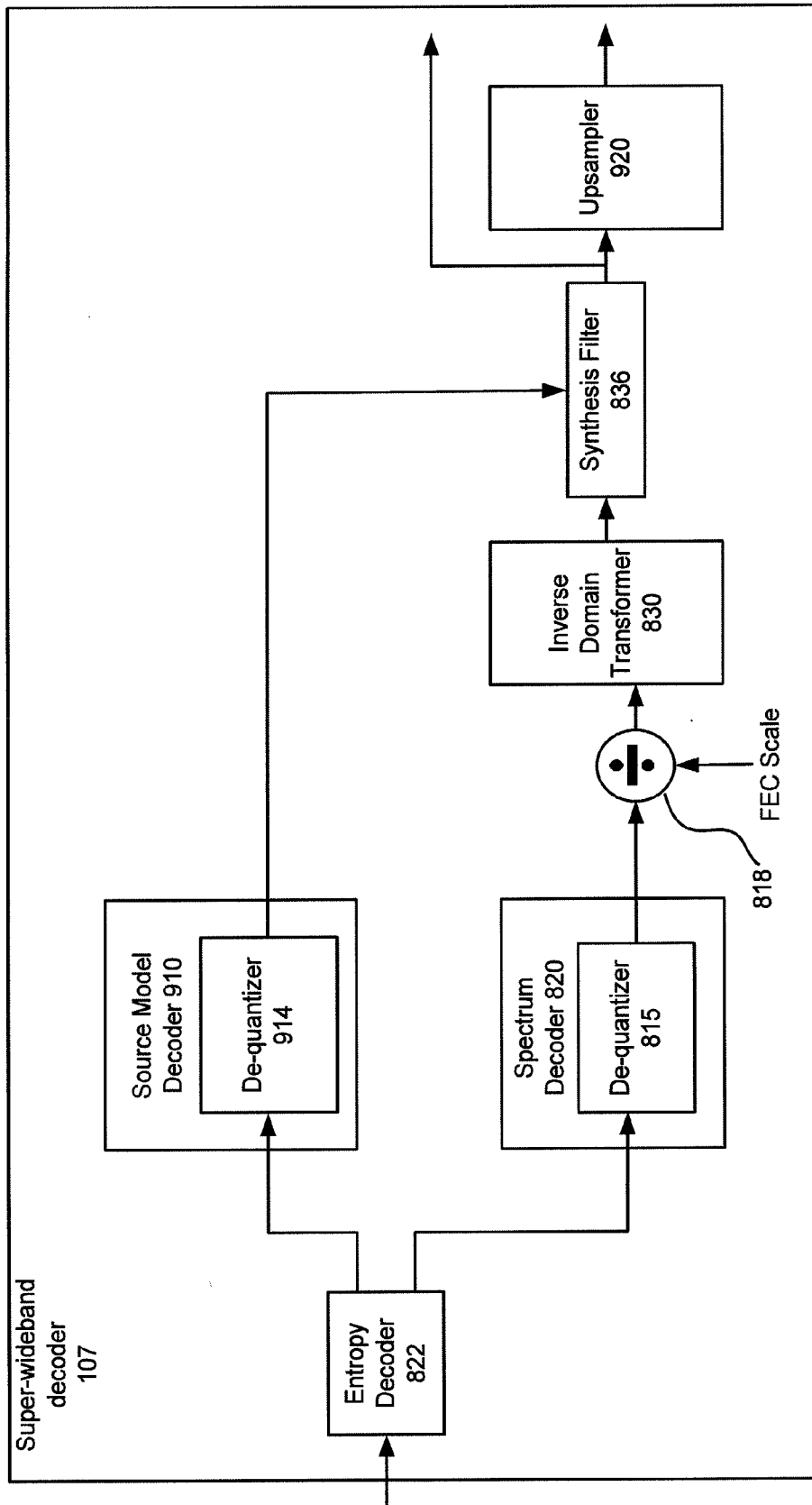


FIG. 9

10/11

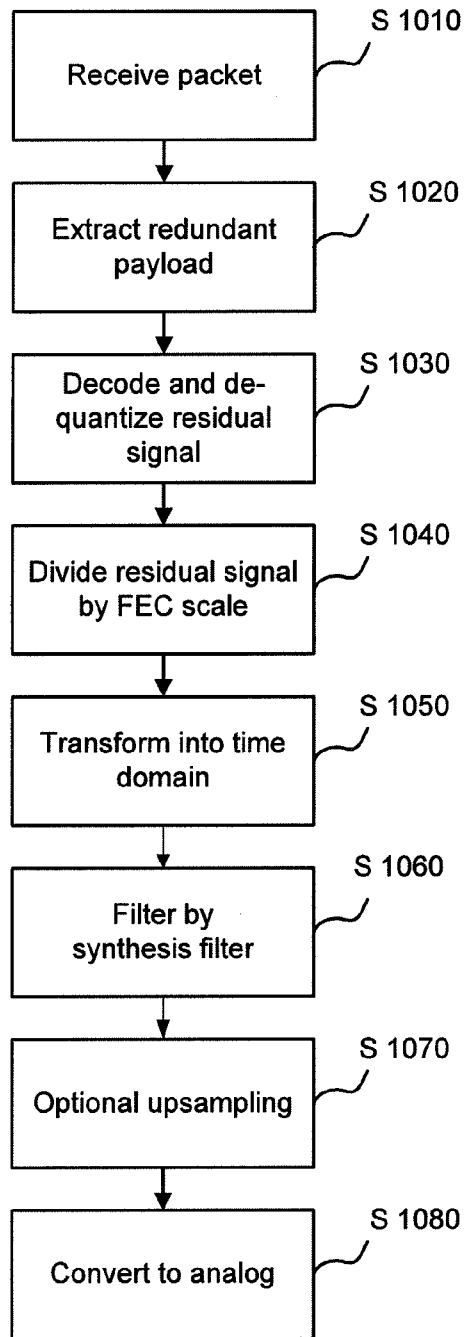


FIG. 10

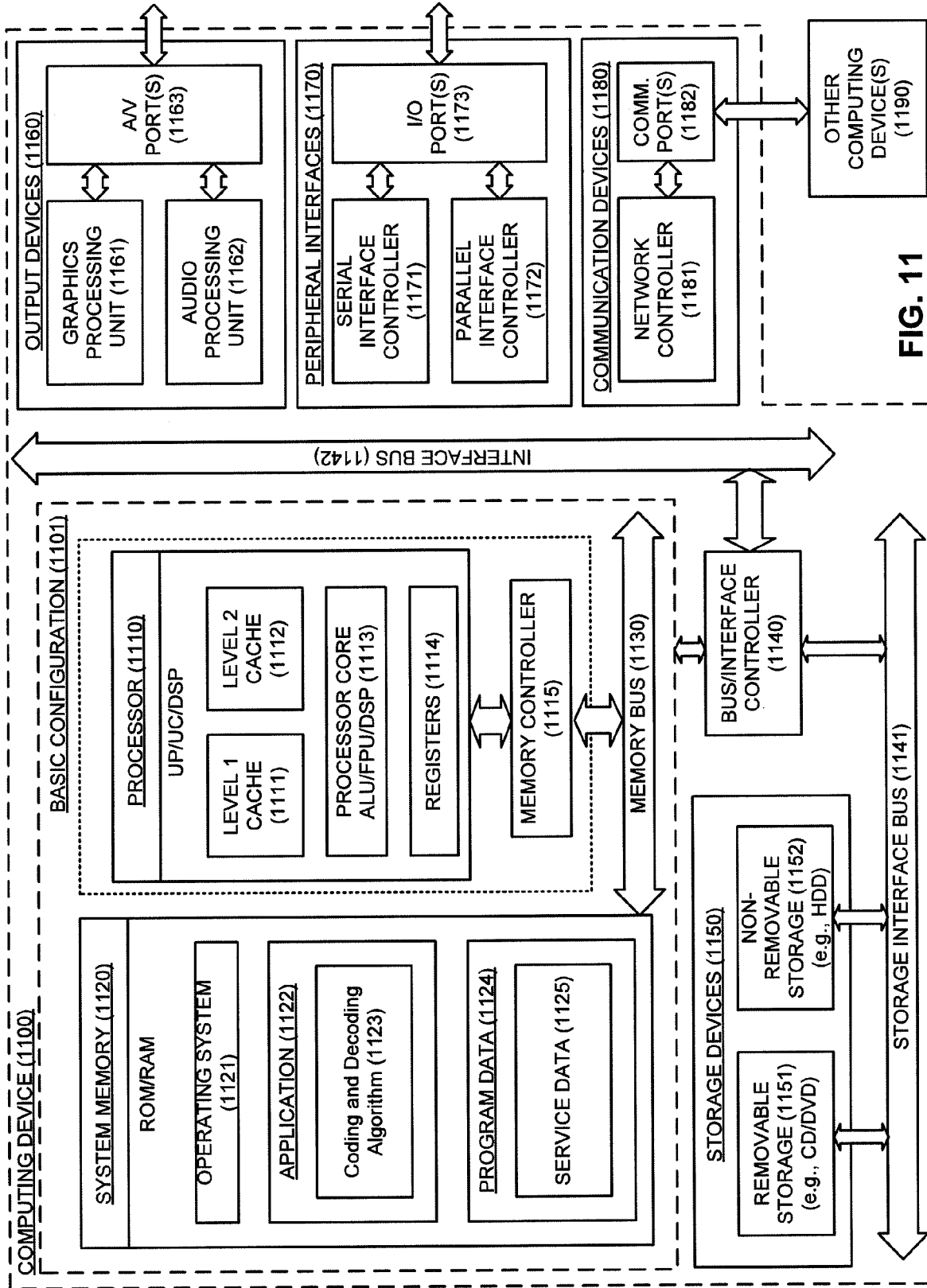


FIG. 11

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2011/037336

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L19/08 G10L19/02 G10L19/00
ADD.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
G10L H04L H03M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SJOBERG M WESTERLUND ERICSSON A LAKANIEMI S WENGER NOKIA J: "RTP Payload Format for the Extended Adaptive Multi-Rate Wideband (AMR-WB+) Audio Codec; rfc4352.txt", 5. JCT-VC MEETING; 96. MPEG MEETING; 16-3-2011 - 23-3-2011; GENEVA;(JOINT COLLABORATIVE TEAM ON VIDEO CODING OF ISO/IEC JTC1/SC29/WG11AND ITU-T SG.16); URL: HTTP://WFTP3.ITU.INT/AV-ARCH/JC TVC-SITE/, INTERNET ENGINEERING TASK FORCE, IETF, CH, 1 January 2006 (2006-01-01), XP015044785, ISSN: 0000-0003	1,4,10, 18,20
Y	page 3 page 9, paragraph 3.6.1 -/--	2,4,11, 14,15,20

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search 14 October 2011	Date of mailing of the international search report 21/10/2011
--	--

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Burchett, Stefanie
--	--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2011/037336

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	& MAKINEN J ET AL: "AMR-WB+: a New Audio Coding Standard for 3rd Generation Mobile Audio Services", 2005 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING - 18-23 MARCH 2005 - PHILADELPHIA, PA, USA, IEEE, PISCATAWAY, NJ, vol. 2, 18 March 2005 (2005-03-18), pages 1109-1112, XP010790838, DOI: 10.1109/ICASSP.2005.1415603 ISBN: 978-0-7803-8874-1	1,4,10, 18,20
Y	pages II-1110, right-hand column, paragraph 3. - pages II-1111, right-hand column, line 3	2,4,11, 14,15,20
Y	----- WO 97/31367 A1 (AT & T CORP [US]; CHEN JUIN HWEY [US]) 28 August 1997 (1997-08-28) page 6, line 13 - page 9, line 25	2,4,11, 20
Y	----- JOHANSSON I ET AL: "Bandwidth efficient amr operation for voip", SPEECH CODING, 2002, IEEE WORKSHOP PROCEEDINGS. OCT. 6-9, 2002, PISCATAWAY, NJ, USA, IEEE, 6 October 2002 (2002-10-06), pages 150-152, XP010647243, ISBN: 978-0-7803-7549-9	14,15
A	page 150, right-hand column, paragraph 2.	5-8
A	----- US 2008/040105 A1 (WANG TIAN [US] ET AL) 14 February 2008 (2008-02-14) paragraph [0059] paragraph [0060] - paragraph [0062] paragraph [0068] - paragraph [0069]	1-20
A	----- KAMARUZZAMAN MD ET AL: "Embedded Speech Codec Based on Speex", AES CONVENTION 116; MAY 2004, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, 1 May 2004 (2004-05-01), XP040506792, page 2, right-hand column, paragraph 2. - page 4, paragraph 2.3	1-20

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2011/037336

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9731367	A1 28-08-1997	CA 2219358 A1 EP 0954851 A1 JP H11504733 A	28-08-1997 10-11-1999 27-04-1999
US 2008040105	A1 14-02-2008	US 2008040121 A1	14-02-2008