SYSTEMS AND METHODS FOR MODIFYING A WINDOW WITH A FRAME ASSOCIATED WITH AN AUDIO SIGNAL

CLAIM OF PRIORITY UNDER 35 U.S.C. § 119

[0001] This present Application for Patent claims priority to Provisional Application No. 60/834,674 entitled "Windowing for Perfect Reconstruction in MDCT with Less than 50 % Frame Overlap" filed July 31, 2006, and assigned to the assignee hereof and hereby expressly incorporated by reference herein.

TECHNICAL FIELD

[0002] The present systems and methods relates generally to speech processing technology. More specifically, the present systems and methods relate to modifying a window with a frame associated with an audio signal.

BACKGROUND

[0003] Transmission of voice by digital techniques has become widespread, particularly in long distance, digital radio telephone applications, video messaging using computers, etc. This, in turn, has created interest in determining the least amount of information that can be sent over a channel while maintaining the perceived quality of the reconstructed speech. Devices for compressing speech find use in many fields of telecommunications. One example of telecommunications is wireless communications. Another example is communications over a computer network, such as the Internet. The field of communications has many applications including, e.g., computers, laptops, personal digital assistants (PDAs), cordless telephones, pagers, wireless local loops, wireless telephony such as cellular and portable communication system (PCS) telephone systems, mobile Internet Protocol (IP) telephony and satellite communication systems.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] Figure 1 illustrates one configuration of a wireless communication system;

[0005] Figure 2 is a block diagram illustrating one configuration of a computing environment;

[0006] Figure 3 is a block diagram illustrating one configuration of a signal transmission environment;

[0007] Figure 4A is a flow diagram illustrating one configuration of a method for modifying a window with a frame associated with an audio signal;

[0008] Figure 4B is a block diagram illustrating a configuration of an encoder for modifying the window with the frame associated with the audio signal and a decoder;

[0009] Figure 5 is a flow diagram illustrating one configuration of a method for reconstructing an encoded frame of an audio signal;

[0010] Figure 6 is a block diagram illustrating one configuration of a multi-mode encoder communicating with a multi-mode decoder;

[0011] Figure 7 is a flow diagram illustrating one example of an audio signal encoding method;

[0012] Figure 8 is a block diagram illustrating one configuration of a plurality of frames after a window function has been applied to each frame;

[0013] Figure 9 is a flow diagram illustrating one configuration of a method for applying a window function to a frame associated with a non-speech signal;

[0014] Figure 10 is a flow diagram illustrating one configuration of a method for reconstructing a frame that has been modified by the window function; and

[0015] Figure 11 is a block diagram of certain components in one configuration of a communication/computing device.

DETAILED DESCRIPTION

[0016] A method for modifying a window with a frame associated with an audio signal is described. A signal is received. The signal is partitioned into a plurality of frames. A determination is made if a frame within the plurality of frames is associated with a non-speech signal. A modified discrete cosine transform (MDCT) window function is applied to the frame to generate a first zero pad region and a second zero pad region if it was determined that the frame is associated with a non-speech signal. The frame is encoded.

[0017] An apparatus for modifying a window with a frame associated with an audio signal is also described. The apparatus includes a processor and memory in electronic communication with the processor. Instructions are stored in the memory. The instructions are executable to: receive a signal; partition the signal into a plurality of

frames; determine if a frame within the plurality of frames is associated with a non-speech signal; apply a modified discrete cosine transform (MDCT) window function to the frame to generate a first zero pad region and a second zero pad region if it was determined that the frame is associated with a non-speech signal; and encode the frame.

[0018] A system that is configured to modify a window with a frame associated with an audio signal is also described. The system includes a means for processing and a means for receiving a signal. The system also includes a means for partitioning the signal into a plurality of frames and a means for determining if a frame within the plurality of frames is associated with a non-speech signal. The system further includes a means for applying a modified discrete cosine transform (MDCT) window function to the frame to generate a first zero pad region and a second zero pad region if it was determined that the frame is associated with a non-speech signal and a means for encoding the frame.

[0019] A computer-readable medium configured to store a set of instructions is also described. The instructions are executable to: receive a signal; partition the signal into a plurality of frames; determine if a frame within the plurality of frames is associated with a non-speech signal; apply a modified discrete cosine transform (MDCT) window function to the frame to generate a first zero pad region and a second zero pad region if it was determined that the frame is associated with a non-speech signal; and encode the frame.

[0020] A method for selecting a window function to be used in calculating a modified discrete cosine transform (MDCT) of a frame is also described. An algorithm for selecting a window function to be used in calculating an MDCT of a frame is provided. The selected window function is applied to the frame. The frame is encoded with an MDCT coding mode based on constraints imposed on the MDCT coding mode by additional coding modes, wherein the constraints comprise a length of the frame, a look ahead length and a delay.

[0021] A method for reconstructing an encoded frame of an audio signal is also described. A packet is received. The packet is disassembled to retrieve an encoded frame. Samples of the frame that are located between a first zero pad region and a first region are synthesized. An overlap region of a first length is added with a look-ahead length of a previous frame. A look-ahead of the first length of the frame is stored. A reconstructed frame is outputted.

[0022] Various configurations of the systems and methods are now described with reference to the Figures, where like reference numbers indicate identical or functionally similar elements. The features of the present systems and methods, as generally described and illustrated in the Figures herein, could be arranged and designed in a wide variety of different configurations. Thus, the detailed description below is not intended to limit the scope of the systems and methods, as claimed, but is merely representative of the configurations of the systems and methods.

[0023] Many features of the configurations disclosed herein may be implemented as computer software, electronic hardware, or combinations of both. To clearly illustrate this interchangeability of hardware and software, various components will be described generally in terms of their functionality. Whether such functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. Skilled artisans may implement the described functionality in varying ways for each particular application, but such implementation decisions should not be interpreted as causing a departure from the scope of the present systems and methods.

[0024] Where the described functionality is implemented as computer software, such software may include any type of computer instruction or computer executable code located within a memory device and/or transmitted as electronic signals over a system bus or network. Software that implements the functionality associated with components described herein may comprise a single instruction, or many instructions, and may be distributed over several different code segments, among different programs, and across several memory devices.

[0025] As used herein, the terms "a configuration," "configuration," "configurations," "the configuration," "the configurations," "one or more configurations," "some configurations," "certain configurations," "one configuration," "another configuration" and the like mean "one or more (but not necessarily all) configurations of the disclosed systems and methods," unless expressly specified otherwise.

[0026] The term "determining" (and grammatical variants thereof) is used in an extremely broad sense. The term "determining" encompasses a wide variety of actions and therefore "determining" can include calculating, computing, processing, deriving, investigating, looking up (e.g., looking up in a table, a database or another data

structure), ascertaining and the like. Also, "determining" can include receiving (e.g., receiving information), accessing (e.g., accessing data in a memory) and the like. Also, "determining" can include resolving, selecting, choosing, establishing, and the like.

[0027] The phrase "based on" does not mean "based only on," unless expressly specified otherwise. In other words, the phrase "based on" describes both "based only on" and "based at least on." In general, the phrase, “audio signal” may be used to refer to a signal that may be heard. Examples of audio signals may include representing human speech, instrumental and vocal music, tonal sounds, etc.

[0028] Figure 1 illustrates a code-division multiple access (CDMA) wireless telephone system 100 that may include a plurality of mobile stations 102, a plurality of base stations 104, a base station controller (BSC) 106 and a mobile switching center (MSC) 108. The MSC 108 may be configured to interface with a public switch telephone network (PSTN) 110. The MSC 108 may also be configured to interface with the BSC 106. There may be more than one BSC 106 in the system 100. Each base station 104 may include at least one sector (not shown), where each sector may have an omnidirectional antenna or an antenna pointed in a particular direction radially away from the base stations 104. Alternatively, each sector may include two antennas for diversity reception. Each base station 104 may be designed to support a plurality of frequency assignments. The intersection of a sector and a frequency assignment may be referred to as a CDMA channel. The mobile stations 102 may include cellular or portable communication system (PCS) telephones.

[0029] During operation of the cellular telephone system 100, the base stations 104 may receive sets of reverse link signals from sets of mobile stations 102. The mobile stations 102 may be conducting telephone calls or other communications. Each reverse link signal received by a given base station 104 may be processed within that base station 104. The resulting data may be forwarded to the BSC 106. The BSC 106 may provide call resource allocation and mobility management functionality including the orchestration of soft handoffs between base stations 104. The BSC 106 may also route the received data to the MSC 108, which provides additional routing services for interface with the PSTN 110. Similarly, the PSTN 18 may interface with the MSC 108, and the MSC 108 may interface with the BSC 106, which in turn may control the base stations 104 to transmit sets of forward link signals to sets of mobile stations 102.

[0030] Figure 2 depicts one configuration of a computing environment 200 including a source computing device 202, a receiving computing device 204 and a receiving mobile computing device 206. The source computing device 202 may communicate with the receiving computing devices 204, 206 over a network 210. The network 210 may a type of computing network including, but not limited to, the Internet, a local area network (LAN), a campus area network (CAN), a metropolitan area network (MAN), a wide area network (WAN), a ring network, a star network, a token ring network, etc.

[0031] In one configuration, the source computing device 202 may encode and transmit audio signals 212 to the receiving computing devices 204, 206 over the network 210. The audio signals 212 may include speech signals, music signals, tones, background noise signals, etc. As used herein, "speech signals" may refer to signals generated by a human speech system and "non-speech signals" may refer to signals not generated by the human speech system (i.e., music, background noise, etc.). The source computing device 202 may be a mobile phone, a personal digital assistant (PDA), a laptop computer, a personal computer or any other computing device with a processor. The receiving computing device 204 may be a personal computer, a telephone, etc. The receiving mobile computing device 206 may be a mobile phone, a PDA, a laptop computer or any other mobile computing device with a processor.

[0032] Figure 3 depicts a signal transmission environment 300 including an encoder 302, a decoder 304 and a transmission medium 306. The encoder 302 may be implemented within a mobile station 102 or a source computing device 202. The decoder 304 may be implemented in a base station 104, in the mobile station 102, in a receiving computing device 204 or in a receiving mobile computing device 206. The encoder 302 may encode an audio signal s(n) 310, forming an encoded audio signal $s_{enc}(n)$ 312. The encoded audio signal 312 may be transmitted across the transmission medium 306 to the decoder 304. The transmission medium 306 may facilitate the encoder 302 to transmit an encoded audio signal 312 to the decoder wirelessly or it may facilitate the encoder 302 to transmit the encoded signal 312 over a wired connection between the encoder 302 and the decoder 304. The decoder 304 may decode $s_{enc}(n)$ 312, thereby generating a synthesized audio signal $\hat{s}(n)$ 316.

[0033] The term "coding" as used herein may refer generally to methods encompassing both encoding and decoding. Generally, coding systems, methods and

apparatuses seek to minimize the number of bits transmitted via the transmission medium 306 (i.e., minimize the bandwidth of $s_{enc}(n)$ 312) while maintaining acceptable signal reproduction (i.e., s(n) 310 $\approx$ $\hat{s}(n)$ 316). The composition of the encoded audio signal 312 may vary according to the particular audio coding mode utilized by the encoder 302. Various coding modes are described below.

[0034] The components of the encoder 302 and the decoder 304 described below may be implemented as electronic hardware, as computer software, or combinations of both. These components are described below in terms of their functionality. Whether the functionality is implemented as hardware or software may depend upon the particular application and design constraints imposed on the overall system. The transmission medium 306 may represent many different transmission media, including, but not limited to, a land-based communication line, a link between a base station and a satellite, wireless communication between a cellular telephone and a base station, between a cellular telephone and a satellite or communications between computing devices.

[0035] Each party to a communication may transmit data as well as receive data. Each party may utilize an encoder 302 and a decoder 304. However, the signal transmission environment 300 will be described below as including the encoder 302 at one end of the transmission medium 306 and the decoder 304 at the other.

[0036] In one configuration, s(n) 310 may include a digital speech signal obtained during a typical conversation including different vocal sounds and periods of silence. The speech signal s(n) 310 may be partitioned into frames, and each frame may be further partitioned into subframes. These arbitrarily chosen frame/subframe boundaries may be used where some block processing is performed. Operations described as being performed on frames might also be performed on subframes, in this sense; frame and subframe are used interchangeably herein. Also, one or more frame may be included in a window which may illustrate the placement and timing between various frames.

[0037] In another configuration, s(n) 310 may include a non-speech signal, such as a music signal. The non-speech signal may be partitioned into frames. One or more frames may be included in a window which may illustrate the placement and timing between various frames. The selection of the window may depend on coding techniques implemented to encode the signal and delay constraints that may be imposed on the system. The present systems and methods describe a method for selecting a

window shape employed in encoding and decoding non-speech signals with a modified discrete cosine transform (MDCT) and an inverse modified discrete cosine transform (IMDCT) based coding technique in a system that is capable of coding both speech and non-speech signals. The system may impose constraints on how much frame delay and look ahead may be used by the MDCT based coder to enable generation of encoded information at a uniform rate.

[0038] In one configuration, the encoder 302 includes a window formatting module 308 which may format the window which includes frames associated with non-speech signals. The frames included in the formatted window may be encoded and the decoder may reconstruct the coded frames by implementing a frame reconstruction module 314. The frame reconstruction module 314 may synthesize the coded frames such that the frames resemble the pre-coded frames of the speech signal 310.

[0039] Figure 4 is a flow diagram illustrating one configuration of a method 400 for modifying a window with a frame associated with an audio signal. The method 400 may be implemented by the encoder 302. In one configuration, a signal is received 402. The signal may be an audio signal as previously described. The signal may be partitioned 404 into a plurality of frames. A window function may be applied 408 to generate a window and a first zero-pad region and a second zero-pad region may be generated as a part of the window for calculating a modified discrete cosine transform (MDCT). In other words, the value of the beginning and end portions of the window may be zero. In one aspect, the length of the first zero-pad region and the length of the second zero-pad region may be a function of delay constraints of the encoder 302.

[0040] The modified discrete cosine transform (MDCT) function may be used in several audio coding standards to transform pulse-code modulation (PCM) signal samples, or their processed versions, into their equivalent frequency domain representation. The MDCT may be similar to a type IV Discrete Cosine Transform (DCT) with the additional property of frames overlapping one another. In other words, consecutive frames of a signal that are transformed by the MDCT may overlap each other by 50%.

[0041] Additionally, for each frame of 2M samples, the MDCT may produce M transform coefficients. The MDCT may be a critically sampled perfect reconstruction filter bank. In order to provide perfect reconstruction, the MDCT coefficients $X(k)$,

for k = 0, 1,...M, obtained from a frame of signal $x(n)$, for n= 0, 1 , . . . $2M$, may be given by

$$X(k) = \sum_{n=0}^{2M-1} x(n)h_k(n) \quad (1)$$

where

$$h_k(n) = w(n)\sqrt{\frac{2}{M}}\cos\left[\frac{(2n+M+1)(2k+1)\pi}{4M}\right] \quad (2)$$

for k = 0, 1,..., M , and $w(n)$ is a window that may satisfy the Princen-Bradley condition, which states:

$$w^2(n) + w^2(n+M) = 1 \quad (3)$$

[0042] At the decoder, the M coded coefficients may be transformed back to the time domain using an inverse MDCT (IMDCT). If $\hat{X}(k)$, for $k = 0,1,2...M$ are the received MDCT coefficients, then the corresponding IMDCT decoder generates the reconstructed audio signal by first taking the IMDCT of the received coefficients to obtain 2M samples according to

$$\hat{x}(n) = \sum_{k=0}^{M-1} \hat{X}(k)h_k(n) \text{ for } n = 0,1,...,2M-1 \quad (4)$$

where $h_k(n)$ is defined by equation (2), then overlapping and adding the first M samples of the present frame with the M last samples of the previous frame's IMDCT output and first M samples from the next frame's IMDCT output. Thus, if the decoded MDCT coefficients corresponding to the next frame are not available at a given time, only M audio samples of the present frame may be completely reconstructed.

[0043] The MDCT system may utilize a look-ahead of M samples. The MDCT system may include an encoder which obtains the MDCT of either the audio signal or filtered versions of it using a predetermined window and a decoder that includes an IMDCT function that uses the same window that the encoder uses. The MDCT system may also include an overlap and an add module. For example, Figure 4B illustrates a MDCT encoder 401. An input audio signal 403 is received by a preprocessor 405. The preprocessor 405 implements preprocessing, linear predictive coding (LPC) filtering and other types of filtering. A processed audio signal 407 is produced from the preprocessor 405. An MDCT function 409 is applied on 2M signal simples that have been appropriately windowed. In one configuration, a quantizer 411 quantizes and

encodes M coefficients 413 and the M coded coefficients are transmitted to an MDCT decoder 429.

[0044] The decoder 429 receives M coded coefficients 413. An IMDCT 415 is applied on the M received coefficients 413 using the same window as in the encoder 401. 2M signal values 417 may be categorized as first M samples selection 423 and last M samples 419 may be saved. The last M samples 419 may further be delayed one frame by a delay 421. The first M samples 423 and the delayed last M samples 419 may be summed by a summer 425. The summed samples may be used to produce a reconstructed M samples 427 of the audio signal.

[0045] Typically, in MDCT systems, 2M signals may be derived from M samples of a present frame and M samples of a future frame. However, if only L samples from the future frame are available, a window may be selected that implements L samples of the future frame.

[0046] In a real-time voice communication system operating over a circuit switched network, the length of the look-ahead samples may be constrained by the maximum allowable encoding delay. It may be assumed that a look-ahead length of L is available. L may be less than or equal to M. Under this condition, it may still be desirable to use the MDCT, with the overlap between consecutive frames being L samples, while preserving the perfect reconstruction property.

[0047] The present systems and methods may be relevant particularly for real time two way communication systems where an encoder is expected to generate information for transmission at a regular interval regardless of the choice of a coding mode. The system may not be capable of tolerating jitter in the generation of such information by the encoder or such a jitter in the generation of such information may not be desired.

[0048] In one configuration, a modified discrete cosine transform (MDCT) function is applied 410 to the frame. Applying the window function may be a step in calculating an MDCT of the frame. In one configuration, the MDCT function processes 2M input samples to generate M coefficients that may then be quantized and transmitted.

[0049] In one configuration, the frame may be encoded 412. In one aspect, the coefficients of the frame may be encoded 412. The frame may be encoded using various encoding modes which will be more fully discussed below. The frame may be formatted 414 into a packet and the packet may be transmitted 416. In one configuration, the packet is transmitted 416 to a decoder.

[0050] Figure 5 is a flow diagram illustrating one configuration of a method 500 for reconstructing an encoded frame of an audio signal. In one configuration, the method 500 may be implemented by the decoder 304. A packet may be received 502. The packet may be received 502 from the encoder 302. The packet may be disassembled 504 in order to retrieve a frame. In one configuration, the frame may be decoded 506. The frame may be reconstructed 508. In one example, the frame reconstruction module 314 reconstructs the frame to resemble the pre-encoded frame of the audio signal. The reconstructed frame may be outputted 510. The outputted frame may be combined with additional outputted frames to reproduce the audio signal.

[0051] Figure 6 is a block diagram illustrating one configuration of a multi-mode encoder 602 communicating with a multi-mode decoder 604 across a communications channel 606. A system that includes the multi-mode encoder 602 and the multi-mode decoder 604 may be an encoding system that includes several different coding schemes to encode different audio signal types. The communication channel 606 may include a radio frequency (RF) interface. The encoder 602 may include an associated decoder (not shown). The encoder 602 and its associated decoder may form a first coder. The decoder 604 may include an associated encoder (not shown). The decoder 604 and its associated encoder may form a second coder.

[0052] The encoder 602 may include an initial parameter calculation module 618, a mode classification module 622, a plurality of encoding modes 624, 626, 628 and a packet formatting module 630. The number of encoding modes 624, 626, 628 is shown as N, which may signify any number of encoding modes 624, 626, 628. For simplicity, three encoding modes 624, 626, 628 are shown, with a dotted line indicating the existence of other encoding modes.

[0053] The decoder 604 may include a packet disassembler module 632, a plurality of decoding modes 634, 636, 638, a frame reconstruction module 640 and a post filter 642. The number of decoding modes 634, 636, 638 is shown as N, which may signify any number of decoding modes 634, 636, 638. For simplicity, three decoding modes 634, 636, 638 are shown, with a dotted line indicating the existence of other decoding modes.

[0054] An audio signal, s(n) 610, may be provided to the initial parameter calculation module 618 and the mode classification module 622. The signal 610 may be divided into blocks of samples referred to as frames. The value n may designate the

frame number or the value n may designate a sample number in a frame. In an alternate configuration, a linear prediction (LP) residual error signal may be used in place of the audio signal 610. The LP residual error signal may be used by speech coders such as a code excited linear prediction (CELP) coder.

[0055] The initial parameter calculation module 618 may derive various parameters based on the current frame. In one aspect, these parameters include at least one of the following: linear predictive coding (LPC) filter coefficients, line spectral pair (LSP) coefficients, normalized autocorrelation functions (NACFs), open-loop lag, zero crossing rates, band energies, and the formant residual signal. In another aspect, the initial parameter calculation module 618 may preprocess the signal 610 by filtering the signal 610, calculating pitch, etc.

[0056] The initial parameter calculation module 618 may be coupled to the mode classification module 622. The mode classification module 622 may dynamically switch between the encoding modes 624, 626, 628. The initial parameter calculation module 618 may provide parameters to the mode classification module 622 regarding the current frame. The mode classification module 622 may be coupled to dynamically switch between the encoding modes 624, 626, 628 on a frame-by-frame basis in order to select an appropriate encoding mode 624, 626, 628 for the current frame. The mode classification module 622 may select a particular encoding mode 624, 626, 628 for the current frame by comparing the parameters with predefined threshold and/or ceiling values. For example, a frame associated with a non-speech signal may be encoded using MDCT coding schemes. An MDCT coding scheme may receive a frame and apply a specific MDCT window format to the frame. An example of the specific MDCT window format is described below in relation to Figure 8.

[0057] The mode classification module 622 may classify a speech frame as speech or inactive speech (e.g., silence, background noise, or pauses between words). Based upon the periodicity of the frame, the mode classification module 622 may classify speech frames as a particular type of speech, e.g., voiced, unvoiced, or transient.

[0058] Voiced speech may include speech that exhibits a relatively high degree of periodicity. A pitch period may be a component of a speech frame that may be used to analyze and reconstruct the contents of the frame. Unvoiced speech may include consonant sounds. Transient speech frames may include transitions between voiced and

unvoiced speech. Frames that are classified as neither voiced nor unvoiced speech may be classified as transient speech.

[0059] Classifying the frames as either speech or non-speech may allow different encoding modes 624, 626, 628 to be used to encode different types of frames, resulting in more efficient use of bandwidth in a shared channel, such as the communication channel 606.

[0060] The mode classification module 622 may select an encoding mode 624, 626, 628 for the current frame based upon the classification of the frame. The various encoding modes 624, 626, 628 may be coupled in parallel. One or more of the encoding modes 624, 626, 628 may be operational at any given time. In one configuration, one encoding mode 624, 626, 628 is selected according to the classification of the current frame.

[0061] The different encoding modes 624, 626, 628 may operate according to different coding bit rates, different coding schemes, or different combinations of coding bit rate and coding scheme. The different encoding modes 624, 626, 628 may also apply a different window function to a frame. The various coding rates used may be full rate, half rate, quarter rate, and/or eighth rate. The various coding modes 624, 626, 628 used may be MDCT coding, code excited linear prediction (CELP) coding, prototype pitch period (PPP) coding (or waveform interpolation (WI) coding), and/or noise excited linear prediction (NELP) coding. Thus, for example, a particular encoding mode 624, 626, 628 may be MDCT coding scheme, another encoding mode may be full rate CELP, another encoding mode 624, 626, 628 may be half rate CELP, another encoding mode 624, 626, 628 may be full rate PPP, and another encoding mode 624, 626, 628 may be NELP.

[0062] In accordance with an MDCT coding scheme that uses a traditional window to encode, transmit, receive and reconstruct at the decoder M samples of an audio signal, the MDCT coding scheme utilizes 2M samples of the input signal at the encoder. In other words, in addition to M samples of the present frame of the audio signal, the encoder may wait for an additional M samples to be collected before the encoding may begin. In a multimode coding system where the MDCT coding scheme co-exists with other coding modes such as CELP, the use of traditional window formats for the MDCT calculation may affect the overall frame size and look ahead lengths of the entire coding system. The present systems and methods provide the design and selection of window

formats for MDCT calculations for any given frame size and look ahead length so that the MDCT coding scheme does not pose constraints on the multimode coding system.

[0063] In accordance with a CELP encoding mode a linear predictive vocal tract model may be excited with a quantized version of the LP residual signal. In CELP encoding mode, the current frame may be quantized. The CELP encoding mode may be used to encode frames classified as transient speech.

[0064] In accordance with a NELP encoding mode a filtered, pseudo-random noise signal may be used to model the LP residual signal. The NELP encoding mode may be a relatively simple technique that achieves a low bit rate. The NELP encoding mode may be used to encode frames classified as unvoiced speech.

[0065] In accordance with a PPP encoding mode a subset of the pitch periods within each frame may be encoded. The remaining periods of the speech signal may be reconstructed by interpolating between these prototype periods. In a time-domain implementation of PPP coding, a first set of parameters may be calculated that describes how to modify a previous prototype period to approximate the current prototype period. One or more codevectors may be selected which, when summed, approximate the difference between the current prototype period and the modified previous prototype period. A second set of parameters describes these selected codevectors. In a frequency-domain implementation of PPP coding, a set of parameters may be calculated to describe amplitude and phase spectra of the prototype. In accordance with the implementation of PPP coding, the decoder 604 may synthesize an output audio signal 616 by reconstructing a current prototype based upon the sets of parameters describing the amplitude and phase. The speech signal may be interpolated over the region between the current reconstructed prototype period and a previous reconstructed prototype period. The prototype may include a portion of the current frame that will be linearly interpolated with prototypes from previous frames that were similarly positioned within the frame in order to reconstruct the audio signal 610 or the LP residual signal at the decoder 604 (i.e., a past prototype period is used as a predictor of the current prototype period).

[0066] Coding the prototype period rather than the entire frame may reduce the coding bit rate. Frames classified as voiced speech may be coded with a PPP encoding mode. By exploiting the periodicity of the voiced speech, the PPP encoding mode may achieve a lower bit rate than the CELP encoding mode.

[0067] The selected encoding mode 624, 626, 628 may be coupled to the packet formatting module 630. The selected encoding mode 624, 626, 628 may encode, or quantize, the current frame and provide the quantized frame parameters 612 to the packet formatting module 630. In one configuration, the quantized frame parameters are the encoded coefficients produced from the MDCT coding scheme. The packet formatting module 630 may assemble the quantized frame parameters 612 into a formatted packet 613. The packet formatting module 630 may provide the formatted packet 613 to a receiver (not shown) over a communications channel 606. The receiver may receive, demodulate, and digitize the formatted packet 613, and provide the packet 613 to the decoder 604.

[0068] In the decoder 604, the packet disassembler module 632 may receive the packet 613 from the receiver. The packet disassembler module 632 may unpack the packet 613 in order to retrieve the encoded frame. The packet disassembler module 632 may also be configured to dynamically switch between the decoding modes 634, 636, 638 on a packet-by-packet basis. The number of decoding modes 634, 636, 638 may be the same as the number of encoding modes 624, 626, 628. Each numbered encoding mode 624, 626, 628 may be associated with a respective similarly numbered decoding mode 634, 636, 638 configured to employ the same coding bit rate and coding scheme.

[0069] If the packet disassembler module 632 detects the packet 613, the packet 613 is disassembled and provided to the pertinent decoding mode 634, 636, 638. The pertinent decoding mode 634, 636, 638 may implement MDCT, CELP, PPP or NELP decoding techniques based on the frame within the packet 613. If the packet disassembler module 632 does not detect a packet, a packet loss is declared and an erasure decoder (not shown) may perform frame erasure processing. The parallel array of decoding modes 634, 636, 638 may be coupled to the frame reconstruction module 640. The frame reconstruction module 640 may reconstruct, or synthesize, the frame, outputting a synthesized frame. The synthesized frame may be combined with other synthesized frames to produce a synthesized audio signal, $\hat{s}(n)$ 616, which resembles the input audio signal, s(n) 610.

[0070] Figure 7 is a flow diagram illustrating one example of an audio signal encoding method 700. Initial parameters of a current frame may be calculated 702. In one configuration, the initial parameter calculation module 618 calculates 702 the parameters. For non-speech frames, the parameters may include one or more

coefficients to indicate the frame is a non-speech frame. Speech frames may include parameters of one or more of the following: linear predictive coding (LPC) filter coefficients, line spectral pairs (LSPs) coefficients, the normalized autocorrelation functions (NACFs), the open loop lag, band energies, the zero crossing rate, and the formant residual signal. Non-speech frames may also include parameters such as linear predictive coding (LPC) filter coefficients.

[0071] The current frame may be classified 704 as a speech frame or a non-speech frame. As previously mentioned, a speech frame may be associated with a speech signal and a non-speech frame may be associated with a non-speech signal (i.e. a music signal). An encoder/decoder mode may be selected 710 based on the frame classification made in steps 702 and 704. The various encoder/decoder modes may be connected in parallel, as shown in Figure 6. The different encoder/decoder modes operate according to different coding schemes. Certain modes may be more effective at coding portions of the audio signal s(n) 610 exhibiting certain properties.

[0072] As previously explained, the MDCT coding scheme may be chosen to code frames classified as non-speech frames, such as music. The CELP mode may be chosen to code frames classified as transient speech. The PPP mode may be chosen to code frames classified as voiced speech. The NELP mode may be chosen to code frames classified as unvoiced speech. The same coding technique may frequently be operated at different bit rates, with varying levels of performance. The different encoder/decoder modes in Figure 6 may represent different coding techniques, or the same coding technique operating at different bit rates, or combinations of the above. The selected encoder mode 710 may apply an appropriate window function to the frame. For example, a specific MDCT window function of the present systems and methods may be applied if the selected encoding mode is an MDCT coding scheme. Alternatively, a window function associated with a CELP coding scheme may be applied to the frame if the selected encoding mode is a CELP coding scheme. The selected encoder mode may encode 712 the current frame and format 714 the encoded frame into a packet. The packet may be transmitted 716 to a decoder.

[0073] Figure 8 is a block diagram illustrating one configuration of a plurality of frames 802, 804, 806 after a specific MDCT window function has been applied to each frame. In one configuration, a previous frame 802, a current frame 804 and a future frame 806 may each be classified as non-speech frames. The length 820 of the current

frame 804 may be represented by 2M. The lengths of the previous frame 802 and the future frame 806 may also be 2M. The current frame 804 may include a first zero pad region 810 and a second zero pad region 818. In other words, the values of the coefficients in the first and second zero-pad regions 810, 818 may be zero.

[0074] In one configuration, the current frame 804 also includes an overlap length 812 and a look-ahead length 816. The overlap and look-ahead lengths 812, 816 may be represented as L. The overlap length 812 may overlap the previous frame 802 look-ahead length. In one configuration, the value L is less than the value M. In another configuration, the value L is equal to the value M. The current frame may also include a unity length 814 in which each value of the frame in this length 814 is unity. As illustrated, the future frame 806 may begin at a halfway point 808 of the current frame 804. In other words, the future frame 806 may begin at a length M of the current frame 804. Similarly, the previous frame 802 may end at the halfway point 808 of the current frame 804. As such, there exists a 50% overlap of the previous frame 802 and the future frame 806 on the current frame 804.

[0075] The specific MDCT window function may facilitate a perfect reconstruction of an audio signal at a decoder if the quantizer/MDCT coefficient module faithfully reconstructs the MDCT coefficients at the decoder. In one configuration, the quantizer/MDCT coefficient encoding module may not faithfully reconstruct the MDCT coefficients at the decoder. In this case, reconstruction fidelity of the decoder may depend on the ability of the quantizer/MDCT coefficient encoding module to reconstruct the coefficients faithfully. Applying the MDCT window to a current frame may provide perfect reconstruction of the current frame if it is overlapped by 50% by both a previous frame and a future frame. In addition, the MDCT window may provide perfect reconstruction if a Princen-Bradley condition is satisfied. As previously mentioned, the Princen-Bradley condition may be expressed as:

$$w^2(n) + w^2(n+M) = 1 \tag{3}$$

where *w(n)* may represent the MDCT window illustrated in Figure 8. The condition expressed by equation (3) may imply that a point on a frame 802, 804, 806 added to a corresponding point on different frame 802, 804, 806 will provide a value of unity. For example, a point of the previous frame 802 in the halfway length 808 added to a

corresponding point of the current frame 804 in the halfway length 808 yields a value of unity.

[0076] Figure 9 is a flow diagram illustrating one configuration of a method 900 for applying an MDCT window function to a frame associated with a non-speech signal, such as the present frame 804 described in Figure 8. The process of applying the MDCT window function may be a step in calculating an MDCT. In other words, a perfect reconstruction MDCT may not be applied without using a window that satisfies the conditions of an overlap of 50% between two consecutive windows and the Princen-Bradley condition previously explained. The window function described in the method 900 may be implemented as a part of applying the MDCT function to a frame. In one example, M samples from the present frame 804 may be available as well as L look-ahead samples. L may be an arbitrary value.

[0077] A first zero pad region of (M-L)/2 samples of the present frame 804 may be generated 902. As previously explained, a zero pad may imply that the coefficients of the samples in the first zero pad region 810 may be zero. In one configuration, an overlap length of L samples of the present frame 804 may be provided 904. The overlap length of L samples of the present frame may be overlapped and added 906 with the previous frame 802 reconstructed look-ahead length. The first zero pad region and the overlap length of the present frame 804 may overlap the previous frame 802 by 50%. In one configuration, (M-L) samples of the present frame may be provided 908. L samples of look-ahead for the present frame may also be provided 910. The L samples of look-ahead may overlap the future frame 806. A second zero pad region of (M-L)/2 samples of the present frame may be generated. In one configuration, the L samples of look-ahead and the second zero pad region of the present frame 804 may overlap the future frame 806 by 50%. A frame which has been applied the method 900 may satisfy the Princen-Bradley condition as previously described.

[0078] Figure 10 is a flow diagram illustrating one configuration of a method 1000 for reconstructing a frame that has been modified by the MDCT window function. In one configuration, the method 1000 is implemented by the frame reconstruction module 314. Samples of the present frame 804 may be synthesized 1002 beginning at the end of a first zero pad region 812 to the end of an (M-L) region 814. An overlap region of L samples of the present frame 804 may be added 1004 with a look-ahead length of the previous frame 802. In one configuration, the look-ahead of L samples 816 of the

present frame 804 may be stored 1006 beginning at the end of the (M-L) region 814 to the beginning of a second zero pad region 818. In one example, the look-ahead of L samples 816 may be stored in a memory component of the decoder 304. In one configuration, M samples may be outputted 1008. The outputted M samples may be combined with additional samples to reconstruct the present frame 804.

[0079] Figure 11 illustrates various components that may be utilized in a communication/computing device 1108 in accordance with the systems and methods described herein. The communication/computing device 1108 may include a processor 1102 which controls operation of the device 1108. The processor 1102 may also be referred to as a CPU. Memory 1104, which may include both read-only memory (ROM) and random access memory (RAM), provides instructions and data to the processor 1102. A portion of the memory 1104 may also include non-volatile random access memory (NVRAM).

[0080] The device 1108 may also include a housing 1122 that contains a transmitter 1110 and a receiver 1112 to allow transmission and reception of data between the access terminal 1108 and a remote location. The transmitter 1110 and receiver 1112 may be combined into a transceiver 1120. An antenna 1118 is attached to the housing 1122 and electrically coupled to the transceiver 1120. The transmitter 1110, receiver 1112, transceiver 1120, and antenna 1118 may be used in a communications device 1108 configuration.

[0081] The device 1108 also includes a signal detector 1106 used to detect and quantify the level of signals received by the transceiver 1120. The signal detector 1106 detects such signals as total energy, pilot energy per pseudonoise (PN) chips, power spectral density, and other signals.

[0082] A state changer 1114 of the communications device 1108 controls the state of the communication/computing device 1108 based on a current state and additional signals received by the transceiver 1120 and detected by the signal detector 1106. The device 1108 may be capable of operating in any one of a number of states.

[0083] The communication/computing device 1108 also includes a system determinator 1124 used to control the device 1108 and determine which service provider system the device 1108 should transfer to when it determines the current service provider system is inadequate.

[0084] The various components of the communication/computing device 1108 are coupled together by a bus system 1126 which may include a power bus, a control signal bus, and a status signal bus in addition to a data bus. However, for the sake of clarity, the various busses are illustrated in Figure 11 as the bus system 1126. The communication/computing device 1108 may also include a digital signal processor (DSP) 1116 for use in processing signals.

[0085] Information and signals may be represented using any of a variety of different technologies and techniques. For example, data, instructions, commands, information, signals, bits, symbols, and chips that may be referenced throughout the above description may be represented by voltages, currents, electromagnetic waves, magnetic fields or particles, optical fields or particles, or any combination thereof.

[0086] The various illustrative logical blocks, modules, circuits, and algorithm steps described in connection with the configurations disclosed herein may be implemented as electronic hardware, computer software, or combinations of both. To clearly illustrate this interchangeability of hardware and software, various illustrative components, blocks, modules, circuits, and steps have been described above generally in terms of their functionality. Whether such functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. Skilled artisans may implement the described functionality in varying ways for each particular application, but such implementation decisions should not be interpreted as causing a departure from the scope of the present systems and methods.

[0087] The various illustrative logical blocks, modules, and circuits described in connection with the configurations disclosed herein may be implemented or performed with a general purpose processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array signal (FPGA) or other programmable logic device, discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described herein. A general purpose processor may be a microprocessor, but in the alternative, the processor may be any processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration.

[0088] The steps of a method or algorithm described in connection with the configurations disclosed herein may be embodied directly in hardware, in a software module executed by a processor, or in a combination of the two. A software module may reside in RAM memory, flash memory, ROM memory, erasable programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), registers, hard disk, a removable disk, a compact disc read-only memory (CD-ROM), or any other form of storage medium known in the art. A storage medium may be coupled to the processor such that the processor can read information from, and write information to, the storage medium. In the alternative, the storage medium may be integral to the processor. The processor and the storage medium may reside in an ASIC. The ASIC may reside in a user terminal. In the alternative, the processor and the storage medium may reside as discrete components in a user terminal.

[0089] The methods disclosed herein comprise one or more steps or actions for achieving the described method. The method steps and/or actions may be interchanged with one another without departing from the scope of the present systems and methods. In other words, unless a specific order of steps or actions is specified for proper operation of the configuration, the order and/or use of specific steps and/or actions may be modified without departing from the scope of the present systems and methods. The methods disclosed herein may be implemented in hardware, software or both. Examples of hardware and memory may include RAM, ROM, EPROM, EEPROM, flash memory, optical disk, registers, hard disk, a removable disk, a CD-ROM or any other types of hardware and memory.

[0090] While specific configurations and applications of the present systems and methods have been illustrated and described, it is to be understood that the systems and methods are not limited to the precise configuration and components disclosed herein. Various modifications, changes, and variations which will be apparent to those skilled in the art may be made in the arrangement, operation, and details of the methods and systems disclosed herein without departing from the spirit and scope of the claimed systems and methods.

[0091] What is claimed is:

CLAIMS

1. A method for modifying a window with a frame associated with an audio signal, the method comprising:

receiving a signal;

partitioning the signal into a plurality of frames;

determining if a frame within the plurality of frames is associated with a non-speech signal;

applying a modified discrete cosine transform (MDCT) window function to the frame to generate a first zero pad region and a second zero pad region if it was determined that the frame is associated with a non-speech signal; and

encoding the frame.

2. The method of claim 1, wherein the frame is encoded using an MDCT coding based scheme.

3. The method of claim 1, wherein the frame comprises a length of 2M, wherein M represents a number of samples in the frame.

4. The method of claim 1, wherein the first zero pad region is located at the beginning of the frame.

5. The method of claim 1, wherein the second zero pad region is located at the end of the frame.

6. The method of claim 1, wherein the first zero pad region and the second region comprise a length of (M-L)/2, wherein L is a value that is less than or equal to M, and wherein M is a number of samples in the frame.

7. The method of claim 7, further comprising providing a present overlap region of length L.

8. The method of claim 7, wherein the overlap region of length L overlaps and is added with look-ahead samples associated with a previous frame.

9. The method of claim 1, further comprising providing a look-ahead region of length L, wherein L is less than or equal to M, and wherein M is a number of samples in the frame.

10. The method of claim 9, wherein the look-ahead region of length L overlaps a future overlap region associated with a future frame.

11. The method of claim 1, wherein the first zero pad region and the present overlap region overlap a previous frame by 50%.

12. The method of claim 1, wherein the second zero pad region and the look-ahead region overlap a future frame by 50%.

13. The method of claim 1, wherein a sum of each sample of the frame added with an associated sample from an overlapped frame equals unity.

14. An apparatus for modifying a window with a frame associated with an audio signal comprising:

a processor;

memory in electronic communication with the processor;

instructions stored in the memory, the instructions being executable to:

receive a signal;

partition the signal into a plurality of frames;

determine if a frame within the plurality of frames is associated with a non-speech signal;

apply a modified discrete cosine transform (MDCT) window function to the frame to generate a first zero pad region and a second zero pad region if it was determined that the frame is associated with a non-speech signal; and

encode the frame.

15. The apparatus of claim 14, wherein the frame is encoded using an MDCT coding based scheme.

16. The apparatus of claim 14, wherein the frame comprises a length of samples equal to 2M, wherein M represents a number of samples in the frame.

17. The apparatus of claim 14, wherein the first zero pad region is located at the beginning of the frame.

18. The apparatus of claim 14, wherein the second zero pad region is located at the end of the frame.

19. A system that is configured to modify a window with a frame associated with an audio signal comprising:

means for processing;

means for receiving a signal;

means for partitioning the signal into a plurality of frames;

means for determining if a frame within the plurality of frames is associated with a non-speech signal;

means for applying a modified discrete cosine transform (MDCT) window function to the frame to generate a first zero pad region and a second zero pad region if it was determined that the frame is associated with a non-speech signal; and

means for encoding the frame.

20. A computer-readable medium configured to store a set of instructions executable to:

receive a signal;

partition the signal into a plurality of frames;

determine if a frame within the plurality of frames is associated with a non-speech signal;

apply a modified discrete cosine transform (MDCT) window function to the frame to generate a first zero pad region and a second zero pad region if it was determined that the frame is associated with a non-speech signal; and

encode the frame.

21. A method for selecting a window function to be used in calculating a modified discrete cosine transform (MDCT) of a frame, the method comprising:

providing an algorithm for selecting a window function to be used in calculating an MDCT of a frame;

applying the selected window function to the frame; and

encoding the frame with an MDCT coding mode based on constraints imposed on the MDCT coding mode by additional coding modes, wherein the constraints comprise a length of the frame, a look ahead length and a delay.

22. A method for reconstructing an encoded frame of an audio signal, the method comprising:

receiving a packet;

disassembling the packet to retrieve an encoded frame;

synthesizing samples of the frame that are located between a first zero pad region and a first region;

adding an overlap region of a first length with a look-ahead length of a previous frame;

storing a look-ahead of the first length of the frame; and

outputting a reconstructed frame.