AUDIO CLASSIFICATION BASED ON PERCEPTUAL QUALITY FOR LOW OR MEDIUM BIT RATES

~~[0001] This application claims priority to U.S. Patent Application No. 14/027052 filed on September 13, 2013, entitled "AUDIO CLASSIFICATION BASED ON PERCEPTUAL QUALITY FOR LOW OR MEDIUM BIT RATES", which is a continuation of U.S. Provisional Application No. 61/702342 filed on September 18, 2012, entitled "Improving AUDIO/VOICED Classification Based on Perceptual Quality for Low or Medium Bit Rates", both of which are incorporated herein by reference as if reproduced in its entirety.~~

TECHNICAL FIELD

[0002] The present invention relates generally to audio classification based on perceptual quality for low or medium bit rates.

BACKGROUND

[0003] Audio signals are typically encoded prior to being stored or transmitted in order to achieve audio data compression, which reduces the transmission bandwidth and/or storage requirements of audio data. Audio compression algorithms reduce information redundancy through coding, pattern recognition, linear prediction, and other techniques. Audio compression algorithms can be either lossy or lossless in nature, with lossy compression algorithms achieving greater data compression than lossless compression algorithms.

SUMMARY OF THE INVENTION

[0004] Technical advantages are generally achieved, by embodiments of this disclosure which describe methods and techniques for improving AUDIO/VOICED classification based on perceptual quality for low or medium bit rates.

[0005] In accordance with an embodiment, a method for classifying signals prior to encoding is provided. In this example, the method includes receiving a digital signal comprising audio data. The digital signal is initially classified as an AUDIO signal. The method further includes re-classifying the digital signal as a VOICED signal when one or more periodicity parameters of the digital signal satisfy a criteria, and encoding the digital signal in accordance with a classification of the digital signal. The digital signal is encoded in the frequency-domain when the digital signal is classified as an AUDIO signal. The digital signal is encoded in the time-domain when the digital signal is re-classified as a VOICED signal. An apparatus for performing this method is also provided.

[0006] In accordance with another embodiment, another method for classifying signals prior to encoding is provided. In this example, the method includes receiving a digital signal comprising audio data. The digital signal is initially classified as an AUDIO signal. The method further includes determining normalized pitch correlation values for subframes in the digital signal, determining an average normalized pitch correlation value by averaging the normalized pitch correlation values, and determining pitch differences between subframes in the digital signal by comparing the normalized pitch correlation values associated with the respective subframes. The method further includes re-classifying the digital signal as a VOICED signal when each of the pitch differences is below a first threshold and the averaged normalized pitch correlation value exceeds a second threshold, and encoding the digital signal in accordance with a classification of the digital signal. The digital signal is encoded in the frequency-domain when the digital signal is classified as an AUDIO signal. The digital signal is encoded in the time-domain when the digital signal is classified as a VOICED signal.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIG. 1 illustrates a diagram of an embodiment code-excited linear prediction (CELP) encoder;

[0008] FIG. 2 illustrates a diagram of an embodiment initial decoder;

[0009] FIG. 3 illustrates a diagram of an embodiment encoder;

[0010] FIG. 4 illustrates a diagram of an embodiment decoder;

[0011] FIG. 5 illustrates a graph depicting a pitch period of a digital signal;

One aspect of present invention provides a method for classifying signals prior to encoding, the method comprising:

receiving a digital signal comprising audio data, the digital signal being initially classified as an AUDIO signal;

determining normalized pitch correlation values for subframes in the digital signal;

determining an average normalized pitch correlation value by averaging the normalized pitch correlation values;

determining pitch differences between subframes in the digital signal by comparing the normalized pitch correlation values associated with the respective subframes;

re-classifying the digital signal as a VOICED signal when each of the pitch differences is below a first threshold and the averaged normalized pitch correlation value exceeds a second threshold; and

encoding the digital signal in accordance with a classification of the digital signal, wherein the digital signal is encoded in the frequency-domain when the digital signal is classified as an AUDIO signal, and wherein the digital signal is encoded in the time-domain when the digital signal is classified as a VOICED signal.

In a possible implementation manner of the aspect, the digital signal carries music.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIG. 1 illustrates a diagram of an embodiment code-excited linear prediction (CELP) encoder;

[0008] FIG. 2 illustrates a diagram of an embodiment initial decoder;

[0009] FIG. 3 illustrates a diagram of an embodiment encoder;

[0010] FIG. 4 illustrates a diagram of an embodiment decoder;

[0011] FIG. 5 illustrates a graph depicting a pitch period of a digital signal;

CLAIMS

1. A method for classifying signals prior to encoding, the method comprising:

receiving a digital signal comprising audio data, the digital signal being initially classified as an AUDIO signal;

re-classifying the digital signal as a VOICED signal when one or more periodicity parameters of the digital signal satisfy a criteria; and

encoding the digital signal in accordance with a classification of the digital signal, wherein the digital signal is encoded in the frequency-domain when the digital signal is classified as an AUDIO signal, and wherein the digital signal is encoded in the time-domain when the digital signal is re-classified as a VOICED signal.

2. The method of claim 1, wherein re-classifying the digital signal as a VOICED signal when one or more periodicity parameters of the digital signal satisfy a criteria comprises:

re-classifying the digital signal as a VOICED signal when a coding rate of the digital signal is below a threshold.

3. The method of claim 1, wherein re-classifying the digital signal as a VOICED signal when one or more periodicity parameters of the digital signal satisfy a criteria comprises:

determining pitch differences between subframes in the digital signal; and

reclassifying the digital signal as a VOICED signal when the pitch differences are less than a threshold.

4. The method of claim 3, wherein determining pitch differences between subframes in the digital signal comprises:

determining normalized pitch correlation values for subframes in the digital signal; and

calculating a pitch difference between subframes by finding a difference between the normalized pitch correlation values of the respective subframes.

5. The method of claim 3 or 4, wherein the threshold is three farads.

6. The method of claim 1, wherein re-classifying the digital signal as a VOICED signal when one or more periodicity parameters of the digital signal satisfy a criteria comprises:

determining an average normalized pitch correlation value for subframes in the digital signal; and

reclassifying the digital signal as a VOICED signal when the average normalized pitch correlation value is less than a threshold.

7. The method of claim 6, wherein determining an average normalized pitch correlation value for subframes in the digital signal comprises:

determining a normalized pitch correlation value for each subframe in the digital signal; and

dividing the sum of all normalized pitch correlation values by the number of subframes in the digital signal to obtain the average normalized pitch correlation value.

8. The method of any one of the claims 1 to 7, wherein the digital signal carries non-speech data.

9. The method of any one of the claims 1 to 8, wherein the digital signal carries music data.

10. An audio encoder comprising:

a processor; and

a computer readable storage medium storing programming for execution by the processor, the programming including instructions to:

receive a digital signal comprising audio data, the digital signal being initially classified as an AUDIO signal;

re-classify the digital signal as a VOICED signal when one or more periodicity parameters of the digital signal satisfy a criteria; and

encode the digital signal in accordance with a classification of the digital signal, wherein the digital signal is encoded in the frequency-domain when the digital signal is classified as an AUDIO signal, and wherein the digital signal is encoded in the time-domain when the digital signal is classified as a VOICED signal.

11. The encoder of claim 10, wherein the instructions to re-classify the digital signal as a VOICED signal when one or more periodicity parameters of the digital signal satisfy a criteria include instructions to:

re-classify the digital signal as a VOICED signal when a coding rate of the digital signal is below a threshold.

12. The encoder of claim 10, wherein the instructions to re-classify the digital signal as a VOICED signal when one or more periodicity parameters of the digital signal satisfy a criteria include instructions to:

determine pitch differences between subframes in the digital signal; and

reclassify the digital signal as a VOICED signal when the pitch differences are less than a threshold.

13. The encoder of claim 12, wherein the instructions to determine pitch differences between subframes in the digital signal include instructions to:

determine normalized pitch correlation values for subframes in the digital signal; and

calculate a pitch difference between subframes by finding a difference between the normalized pitch correlation values of the respective subframes.

14. The encoder of claim 12 or 13, wherein the threshold is three farads.

15. The encoder of claim 10, wherein the instructions to re-classify the digital signal as a VOICED signal when one or more periodicity parameters of the digital signal satisfy a criteria include instructions to:

determine an average normalized pitch correlation value for subframes in the digital signal; and

reclassify the digital signal as a VOICED signal when the average normalized pitch correlation value is less than a threshold.

16. The encoder of claim 15, wherein the instructions to determine an average normalized pitch correlation value for subframes in the digital signal include instructions to:

determine a normalized pitch correlation value for each subframe in the digital signal; and

divide the sum of all normalized pitch correlation values by the number of subframes in the digital signal to obtain the average normalized pitch correlation value.

17. The encoder of any one of the claims 10 to 16, wherein the digital signal carries non-speech data.

18. The encoder of any one of the claims 10 to 17, wherein the digital signal carries music data.

19. A method for classifying signals prior to encoding, the method comprising:

receiving a digital signal comprising audio data, the digital signal being initially classified as an AUDIO signal;

determining normalized pitch correlation values for subframes in the digital signal;

determining an average normalized pitch correlation value by averaging the normalized pitch correlation values;

determining pitch differences between subframes in the digital signal by comparing the normalized pitch correlation values associated with the respective subframes;

re-classifying the digital signal as a VOICED signal when each of the pitch differences is

below a first threshold and the averaged normalized pitch correlation value exceeds a second threshold; and

encoding the digital signal in accordance with a classification of the digital signal, wherein the digital signal is encoded in the frequency-domain when the digital signal is classified as an AUDIO signal, and wherein the digital signal is encoded in the time-domain when the digital signal is classified as a VOICED signal.

20. The method of claim 19, wherein the digital signal carries music.

[CLAIMS]

1. A method for encoding signals, the method comprising:
receiving a digital signal comprising audio data, the digital signal being initially classified as an AUDIO signal;
determining whether a classifying condition is satisfied, wherein, the classifying condition comprises: pitch differences between sub-frames in the digital signal are less than a first threshold;
classifying the digital signal as a VOICED signal when the classifying condition is satisfied; and
encoding the digital signal in the frequency-domain when the digital signal is classified as an AUDIO signal, and encoding the digital signal in the time-domain when the digital signal is classified as a VOICED signal;
wherein, the number of the sub-frames is 4, the straight pitch differences comprises the first straight pitch difference dpit1, the second straight pitch difference dpit2, and the third straight pitch difference dpit3, wherein, the dpit1, the dpit2 and the dpit3 are calculated as follows:

$$dpit1 = |P_1 - P_2|$$

$$dpit2 = |P_2 - P_3|$$

$$dpit3 = |P_3 - P_4|,$$

wherein, P1, P2, P3, and P4 are four pitch values corresponding to the sub-frames respectively;
accordingly, and wherein the classifying condition that the straight pitch differences between sub-frames in the digital signal are less than a threshold comprises: all the dpit1, the dpit2 and the dpit3 are less than the first threshold.

2. The method of claim 1, wherein the classifying condition further comprises:
a bit rate of the digital signal is below a threshold.

3. The method of claim 1, wherein the method further comprises:
determining an average normalized pitch correlation value for subframes in the digital signal; and
wherein, the classifying condition further comprising:
the average normalized pitch correlation value for subframes in the digital signal is greater than a threshold.

4. The method of claim 3, wherein determining an average normalized pitch correlation value for subframes in the digital signal comprises:
determining a normalized pitch correlation value for each subframe in the digital signal; and
dividing the sum of all normalized pitch correlation values by the number of subframes in the digital signal to obtain the average normalized pitch correlation value.

5. The method of claim 1, wherein the digital signal carries non-speech data.

6. The method of claim 1, wherein the digital signal carries music data.

7. The method of claim 1, wherein, P1, P2, P3, and P4 are the best pitch values found in a pitch range from a minimum pitch limit PIT_MIN to a maximum pitch limit PIT_MAX for each sub-frame.

8. The method of claim 1, wherein, each of the straight pitch differences is less than the first threshold; the method further comprising:
determining normalized pitch correlation values for sub-frames in the digital signal; and
determining an average normalized pitch correlation value by averaging the normalized pitch correlation values; accordingly, and
wherein the classifying condition further comprises:
the averaged normalized pitch correlation value exceeds a second threshold.

9. The method of claim 8, wherein, determining an average normalized pitch correlation value for sub-frames in the digital signal comprises:
determining a normalized pitch correlation value for each sub-frame in the digital signal; and
dividing the sum of all normalized pitch correlation values by the number of sub-frames in the digital signal to obtain the average normalized pitch correlation value.

10. The method of claim 8, wherein, the method further comprising:
determining a smoothed pitch correlation according to the average normalized pitch correlation value; accordingly, and
wherein the classifying condition further comprises:
the smoothed pitch correlation exceeds a third threshold.

11. The method of claim 10, wherein, the classifying condition further comprises:
a bit rate of the digital signal is below a fourth threshold.

12. An audio encoder comprising:
a processor; and
a computer readable storage medium storing programming for execution by the processor, the programming including instructions to:
receive a digital signal comprising audio data and classifying the digital signal as an AUDIO signal;
determine whether a classifying condition is satisfied, wherein, the classifying condition comprises: straight pitch differences between sub-frames in the digital signal are less than a first threshold;
classify the digital signal as a VOICED signal when the classifying condition is satisfied; and
encode the digital signal in the time-domain if the digital signal is classified as a VOICED signal, and encode the digital signal in the frequency-domain if the digital signal is classified as an AUDIO signal;
wherein, the number of the sub-frames is 4, the s 15 itch differences comprises the first straight pitch difference dpit1, the second straight pitch difference dpit2, the third straight pitch difference dpit3, wherein, the dpit1, the dpit2 and the dpit3 are calculated as follows:

$$dpit1 = |P_1 - P_2|$$

$$dpit2 = |P_2 - P_3|$$

$$dpit3 = |P_3 - P_4|,$$

wherein, P1, P2, P3, and P4 are four pitch values corresponding to the sub-frames respectively;
accordingly, and
wherein the classifying condition that the straight pitch differences between sub-frames in the digital signal are less than a threshold comprises:
all the dpit1, the dpit2 and the dpit3 are less than the first threshold.

13. The encoder of claim 12, wherein the classifying condition further comprises:
a bit rate of the digital signal is below a threshold.

14. The encoder of claim 12, wherein the programming including further instructions to:
determine an average normalized pitch correlation value for subframes in the digital signal; and
wherein the classifying condition further comprises:
the average normalized pitch correlation value is greater than a threshold.

15. The encoder of claim 14, wherein the instructions to determine an average normalized pitch correlation value for subframes in the digital signal include instructions to:
determine a normalized pitch correlation value for each subframe in the digital signal; and
divide the sum of all normalized pitch correlation values by the number of subframes in the digital signal to obtain the average normalized pitch correlation value.

16. The encoder of claim 12, wherein the digital signal carries non-speech data. 17. The encoder of claim 12, wherein the digital signal carries music data.

18. The encoder of claim 17, wherein, P_1, P_2, P_3, and P_4 are the best pitch values found in a pitch range from a minimum pitch limit *PIT_MIN* to a maximum pitch limit *PIT_MAX* for each sub-frame.

19. The encoder of claim 17, wherein, each of the straight pitch differences is less than the first threshold; the programming including further instructions to:
determine normalized pitch correlation values for sub-frames in the digital signal; and
determine an average normalized pitch correlation value by averaging the normalized pitch correlation values; accordingly, and
wherein the classifying condition further comprises:
the averaged normalized pitch correlation value exceeds a second threshold.

20. The encoder of claim 19, wherein the instructions to determine an average normalized pitch correlation value for sub-frames in the digital signal include instructions to:
determine a normalized pitch correlation value for each sub-frame in the digital signal; and
divide the sum of all normalized pitch correlation values by the number of sub-frames in the digital signal to obtain the average normalized pitch correlation value.

21. The encoder of claim 20, wherein, the programming including further instructions to:
determine a smoothed pitch correlation according to the average normalized pitch correlation value; accordingly, and
wherein the classifying condition further comprises:
the smoothed pitch correlation exceeds a third threshold.

22. The encoder of claim 21, wherein, the classifying condition further comprises:
a bit rate of the digital signal is below a fourth threshold.

23. A method for encoding audio signal comprising a plurality of frames by a processor, comprising:
receiving a frame of the audio signal, wherein the frame comprising a plurality of sub-frames; and
encoding the frame in the time-domain when a classifying condition is satisfied, wherein the classifying condition comprises:
the straight pitch differences between the sub-frames are less than a first threshold;
wherein, the number of the sub-frames is 4, the straight pitch differences comprises the first straight pitch difference *dpit1*, the second straight pitch difference *dpit2*, the third straight pitch difference *dpit3*, wherein, the *dpit1*, the *dpit2* and the *dpit3* are calculated as follows:

$$dpit1 = |P_1 - P_2|$$
$$dpit2 = |P_2 - P_3|$$
$$dpit3 = |P_3 - P_4|,$$

wherein, P_1, P_2, P_3, and P_4 are four pitch values corresponding to the sub-frames respectively;
accordingly, and
wherein the classifying condition that the straight pitch differences between sub-frames in the digital signal are less than a threshold comprises:
all the *dpit1*, the *dpit2* and the *dpit*3 are less than the first threshold.

24. The method of claim 23, wherein the method further comprising:
determining an average normalized pitch correlation value for sub-frames in the digital signal; accordingly, and
wherein the classifying condition further comprises:
the average the average normalized pitch correlation value for sub-frames in the digital signal is greater than a

threshold.

25. The method of claim 24, wherein determining an average normalized pitch correlation value for sub-frames in the digital signal comprises:
determining a normalized pitch correlation value for each sub-frame in the digital signal; and
dividing the sum of all normalized pitch correlation values by the number of sub-frames in the digital signal to obtain the average normalized pitch correlation value.

26. The method of claim 23, wherein, $P_1, P_2, P_3,$ and P_4 are the best pitch values found in a pitch range from a minimum pitch limit *PIT_MIN* to a maximum pitch limit *PIT_MAX* for each sub-frame.

27. The method of claim 23, wherein, each of the straight pitch differences is less than the first threshold; the method further comprising:
determining normalized pitch correlation values for sub-frames in the digital signal; and
determining an average normalized pitch correlation value by averaging the normalized pitch correlation values; accordingly, and
wherein the classifying condition further comprises:
the averaged normalized pitch correlation value exceeds a second threshold.

28. The method of claim 27, wherein, determining an average normalized pitch correlation value for sub-frames in the digital signal comprises:
determining a normalized pitch correlation value for each sub-frame in the digital signal; and
dividing the sum of all normalized pitch correlation values by the number of sub-frames in the digital signal to obtain the average normalized pitch correlation value.

29. The method of claim 27, wherein, the method further comprising:
determining a smoothed pitch correlation according to the average normalized pitch correlation value; accordingly, and
wherein the classifying condition further comprises:
the smoothed pitch correlation exceeds a third threshold.

30. The method of claim 29, wherein, the classifying condition further comprises:
a bit rate of the digital signal is below a fourth threshold.