Unvoiced/Voiced Decision for Speech Processing

[1] ~~This application claims priority to U.S. Patent Application No. 14/476,547 filed on September 3, 2014, entitled "Unvoiced/Voiced Decision for Speech Processing", which is a continuation of U.S. Provisional Application No. 61/875,198 filed on September 9, 2013, entitled "Improved Unvoiced/Voiced Decision for Speech Coding / Bandwidth Extension / Speech Enhancement", both of which are incorporated herein by reference as if reproduced in its entirety~~.

TECHNICAL FIELD

[2] The present invention is generally in the field of speech processing, and in particular to Voiced/Unvoiced Decision for speech processing.

BACKGROUND

[3] Speech coding refers to a process that reduces the bit rate of a speech file. Speech coding is an application of data compression of digital audio signals containing speech. Speech coding uses speech-specific parameter estimation using audio signal processing techniques to model the speech signal, combined with generic data compression algorithms to represent the resulting modeled parameters in a compact bitstream. The objective of speech coding is to achieve savings in the required memory storage space, transmission bandwidth and transmission power by reducing the number of bits per sample such that the decoded (decompressed) speech is perceptually indistinguishable from the original speech.

[4] However, speech coders are lossy coders, i.e., the decoded signal is different from the original. Therefore, one of the goals in speech coding is to minimize the distortion (or perceptible loss) at a given bit rate, or minimize the bit rate to reach a given distortion.

SUMMARY

[13] In accordance with an embodiment of the present invention, a method for speech processing comprises determining an unvoicing/voicing parameter reflecting a characteristic of unvoiced/voiceding speech in a current frame of a speech signal comprising a plurality of frames. A smoothed unvoicing/voicing parameter is determined to include information of the unvoicing/voicing parameter in a frame prior to the current frame of the speech signal. A difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter is computed. The method further includes generating an unvoiced/voiced decision point for determining whether the current frame comprises unvoiced speech or voiced speech using the computed difference as a decision parameter.

[14] In an alternative embodiment, a speech processing apparatus comprises a processor, and a computer readable storage medium storing programming for execution by the processor. The programming include instructions to determine an unvoicing/voicing parameter reflecting a characteristic of unvoiced/voicingunvoiced/voiced speech in a current frame of a speech signal comprising a plurality of frames, and determine a smoothed unvoicing/voicing parameter to include information of the unvoicing/voicing parameter in a frame prior to the current frame of the speech signal. The programming further include instructions to compute a difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter, and generate a unvoiced/voiced decision point for determining whether the current frame comprises unvoiced speech or voiced speech using the computed difference as a decision parameter.

[15] In an alternative embodiment, a method for speech processing comprises providing a plurality of frames of a speech signal and determining, for a current frame, a first for a first frequency band from a first energy envelope of the speech signal in the time domain and a

The dots at the end of Equation (~~11~~5) indicate other parameters may be added. When the "absolute" value of $P_{c_unvoicing}$ becomes large, it is likely unvoiced speech signal. A combined voicing parameter could be described as in Equation (6) below.

$$P_{c_voicing} = P_{voicing} \cdot P_{tilt} \cdots \quad (6)$$

The dots at the end of Equation (6) similarly indicate that other parameters may be added. When the "absolute" value of $P_{c_voicing}$ becomes large, it is likely voiced speech signal. Before the "relative" values of $P_{c_unvoicing}$ or $P_{c_voicing}$ are defined, a strongly smoothed parameter of $P_{c_unvoicing}$ or $P_{c_voicing}$ is defined first. For example, the parameter for current frame may be smoothed from a previous frame as described by inequality below in Equation (7).

$$\begin{aligned} &if \quad (P_{c_unvoicing_sm} > P_{c_unvoicing}) \quad \{ \\ &\qquad P_{c_unvoicing_sm} \Leftarrow 0.9\,P_{c_unvoicing_sm} + 0.1\,P_{c_unvoicing} \\ &\} \\ &else \quad \{ \\ &\qquad P_{c_unvoicing_sm} \Leftarrow 0.99\,P_{c_unvoicing_sm} + 0.01\,P_{c_unvoicing} \\ &\} \end{aligned} \quad (7)$$

In Equation (7), $P_{c_unvoicing_sm}$ is a strongly smoothed value of $P_{c_unvoicing}$.

[51] Similarly, the smoothed combined voicing parameter $P_{c_voicing_sm}$ may be determined using the inequality below using Equation (8).

~~unvoiced/voicing~~unvoiced/voiced speech in the current frame is determined (box 1312). In various embodiments, the unvoicing/voicing parameter may include a periodicity parameter, a spectral tilt parameter, or other variants. The method further includes determining a smoothed unvoicing parameter to include information of the unvoicing/voicing parameter in previous frames of the speech signal (box 1314). A difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter is obtained (box 1316). Alternatively, a relative value (e.g., ratio) between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter may be obtained. When deciding whether a current frame is better suited to be handled as an unvoiced/voiced speech, the unvoiced/voiced decision is made using the determined difference as a decision parameter (box 1318).

[111] Referring to Figure 13B, a method for speech processing includes receiving a plurality of frames of a speech signal (box 1320). The embodiment is described using a voicing parameter but equally applies to using an unvoicing parameter. A combined voicing parameter is determined for each frame (box 1322). In one or more embodiments, the combined voicing parameter may be a periodicity parameter and a tilt parameter and a smoothed combined voicing parameter. The smoothed combined voicing parameter may be obtained by smoothing the combined voicing parameter over one or more previous frames of the speech signal. The combined voicing parameter is compared with the smoothed combined voicing parameter (box 1324). The current frame is classified as a VOICED speech signal or an UNVOICED speech signal using the comparison in the decision making (box 1326). The speech signal may be processed, for example, encoded or decoded, in accordance with the determined classification of the speech signal (box 1328).

WHAT IS CLAIMED IS:

1. A method for speech processing, the method comprising:

determining an unvoicing/voicing parameter reflecting a characteristic of unvoiced/voicing speech in a current frame of a speech signal comprising a plurality of frames;

determining a smoothed unvoicing/voicing parameter to include information of the unvoicing/voicing parameter in a frame prior to the current frame of the speech signal;

computing a difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter; and

generating a unvoiced/voiced decision point for determining whether the current frame comprises unvoiced speech or voiced speech using the computed difference as a decision parameter.

2. The method of claim 1, wherein the unvoicing/voicing parameter is a combined parameter reflecting at least two characteristics of unvoiced/voiced speech.

3. The method of claim 2, wherein the combined parameter is a product of a periodicity parameter and a spectral tilt parameter.

4. The method of claim 1, wherein the unvoicing/voicing parameter is an unvoicing parameter ($P_{unvoicing}$) reflecting a characteristic of unvoiced speech, wherein the smoothed unvoicing/voicing parameter is a smoothed unvoicing parameter ($P_{unvoicing_sm}$).

5. The method of claim 4, wherein when the difference between the unvoicing parameter and the smoothed unvoicing parameter is greater than 0.1, determining the unvoiced/voiced decision comprises determining the current frame of the speech signal to be an unvoiced signal,

wherein when the difference between the unvoicing parameter and the smoothed unvoicing parameter is less than 0.05, determining the current frame of the speech signal to be not an unvoiced speech.

6. The method of claim 5, wherein when the difference between the unvoicing parameter and the smoothed unvoicing parameter is between 0.05 to 0.1, determining the unvoiced/voiced decision comprises determining the current frame of the speech signal to have the same speech type as the prior frame.

7. The method of claim 4, wherein the smoothed unvoicing parameter is computed from the unvoicing parameter as follows.

$$
\begin{aligned}
&\textit{if} \quad (P_{unvoicing_sm} > P_{unvoicing}) \quad \{ \\
&\qquad P_{unvoicing_sm} \quad \Leftarrow \quad 0.9\,P_{unvoicing_sm} \quad + \quad 0.1\,P_{unvoicing} \\
&\} \\
&\textit{else} \quad \{ \\
&\qquad P_{unvoicing_sm} \quad \Leftarrow \quad 0.99\,P_{unvoicing_sm} \quad + \quad 0.01\,P_{unvoicing} \\
&\}.
\end{aligned}
$$

8. The method of claim 1, wherein the unvoicing/voicing parameter is a voicing parameter ($P_{voicing}$) reflecting a characteristic of voiced speech, and wherein the smoothed unvoicing/voicing parameter is a smoothed voicing parameter ($P_{voicing_sm}$).

9. The method of claim 8, wherein when the difference between the voicing parameter and the smoothed voicing parameter is greater than 0.1, determining the unvoiced/voiced decision comprises determining the current frame of the speech signal to be an voiced signal, and wherein

when the difference between the voicing parameter and the smoothed voicing parameter is less than 0.05, determining the current frame of the speech signal to be not an voiced speech.

10. The method of claim 8, wherein the smoothed voicing parameter is computed from the voicing parameter as follows.

$$
\begin{aligned}
&if \quad (P_{voicing_sm} > P_{voicing} \quad) \quad \{ \\
&\qquad P_{voicing_sm} \quad \Leftarrow \quad (7/8)\, P_{voicing_sm} \quad + \quad (1/8)\, P_{voicing} \\
&\} \\
&else \quad \{ \\
&\qquad P_{voicing_sm} \quad \Leftarrow \quad (255/256)\, P_{voicing_sm} \quad + \quad (1/256)\, P_{voicing} \\
&\}.
\end{aligned}
$$

11. The method of any of claims 1 to 10, wherein determining an unvoicing/voicing parameter reflecting a characteristic of unvoiced/voicing speech in a current frame comprises determining a first energy envelope of the speech signal in time domain within a first frequency band and a second energy envelope of the speech signal in time domain within a different second frequency band.

12. The method of claim 11, wherein the second frequency band is a higher frequency band than the first frequency band.

13. A speech processing apparatus comprising:

a processor; and

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

determine an unvoicing/voicing parameter reflecting a characteristic of unvoiced/voicing speech in a current frame of a speech signal comprising a plurality of frames,

determine a smoothed unvoicing/voicing parameter to include information of the unvoicing/voicing parameter in a frame prior to the current frame of the speech signal,

compute a difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter, and

generate a unvoiced/voiced decision point for determining whether the current frame comprises unvoiced speech or voiced speech using the computed difference as a decision parameter.

14. The apparatus of claim 13, wherein the unvoicing/voicing parameter is a combined parameter reflecting a product of a periodicity parameter and a spectral tilt parameter.

15. The apparatus of claim 13, wherein when the difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter is greater than 0.1, determining the unvoiced/voiced decision comprises determining the current frame of the speech signal to be an unvoiced/voiced signal, wherein when the difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter is less than 0.05, determining the current frame of the speech signal to be not an unvoiced/voiced speech.

16. The apparatus of claim 13, wherein the unvoicing/voicing parameter is an unvoicing parameter reflecting a characteristic of unvoiced speech, and wherein the smoothed unvoicing/voicing parameter is a smoothed unvoicing parameter.

17. The apparatus of claim 13, wherein the unvoicing/voicing parameter is a voicing parameter reflecting a characteristic of voiced speech, and wherein the smoothed unvoicing/voicing parameter is a smoothed voicing parameter.

18. The apparatus of any one of claims 13 to 17, wherein determining an unvoicing/voicing parameter reflecting a characteristic of unvoiced/voicing speech in a current frame comprises determining a first energy envelope of the speech signal in time domain within a first frequency band and a second energy envelope of the speech signal in time domain within a different second frequency band.

19. The apparatus of claim 18, wherein the second frequency band is a higher frequency band than the first frequency band.

WHAT IS CLAIMED IS:

1. A method for speech processing, the method comprising:

determining an unvoicing/voicing parameter reflecting a characteristic of unvoiced/voiced speech in a current frame of a speech signal comprising a plurality of frames;

determining a smoothed unvoicing/voicing parameter to include information of the unvoicing/voicing parameter in a frame prior to the current frame of the speech signal;

computing a difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter; and

determining whether the current frame comprises unvoiced speech or voiced speech using the computed difference as a decision parameter.

2. The method of claim 1, wherein the unvoicing/voicing parameter is a combined parameter reflecting at least two characteristics of unvoiced/voiced speech.

3. The method of claim 2, wherein the combined parameter is a product of a periodicity parameter and a spectral tilt parameter.

4. The method of any of claims 1 to 3, wherein the unvoicing/voicing parameter is an unvoicing parameter ($P_{unvoicing}$) reflecting a characteristic of unvoiced speech, wherein the smoothed unvoicing/voicing parameter is a smoothed unvoicing parameter ($P_{unvoicing_sm}$).

5. The method of claim 4, wherein when the difference between the unvoicing parameter and the smoothed unvoicing parameter is greater than 0.1, determining the current frame of the speech signal to be an unvoiced signal, wherein when the difference between the unvoicing

parameter and the smoothed unvoicing parameter is less than 0.05, determining the current frame of the speech signal to be not an unvoiced speech.

6. The method of claim 5, wherein when the difference between the unvoicing parameter and the smoothed unvoicing parameter is between 0.05 to 0.1, determining the current frame of the speech signal to have the same speech type as the prior frame.

7. The method of any of claims 4 to 6, wherein the smoothed unvoicing parameter is computed from the unvoicing parameter as follows.

$$
\begin{aligned}
&\textit{if} \quad (P_{unvoicing_sm} > P_{unvoicing}) \quad \{ \\
&\qquad P_{unvoicing_sm} \Leftarrow 0.9\,P_{unvoicing_sm} + 0.1\,P_{unvoicing} \\
&\} \\
&\textit{else} \quad \{ \\
&\qquad P_{unvoicing_sm} \Leftarrow 0.99\,P_{unvoicing_sm} + 0.01\,P_{unvoicing} \\
&\}.
\end{aligned}
$$

8. The method of any of claims 1 to 3, wherein the unvoicing/voicing parameter is a voicing parameter ($P_{voicing}$) reflecting a characteristic of voiced speech, and wherein the smoothed unvoicing/voicing parameter is a smoothed voicing parameter ($P_{voicing_sm}$).

9. The method of claim 8, wherein when the difference between the voicing parameter and the smoothed voicing parameter is greater than 0.1, determining the current frame of the speech signal to be an voiced signal, and wherein when the difference between the voicing parameter and the smoothed voicing parameter is less than 0.05, determining the current frame of the speech signal to be not an voiced speech.

10. The method of claims 8 or 9, wherein the smoothed voicing parameter is computed from the voicing parameter as follows.

$$
\begin{aligned}
&if \quad (P_{voicing_sm} > P_{voicing}) \ \{ \\
&\quad P_{voicing_sm} \Leftarrow (7/8)\, P_{voicing_sm} + (1/8)\, P_{voicing} \\
&\} \\
&else \ \{ \\
&\quad P_{voicing_sm} \Leftarrow (255/256)\, P_{voicing_sm} + (1/256)\, P_{voicing} \\
&\}.
\end{aligned}
$$

11. The method of any of claims 1 to 10, wherein the frame comprises a subframe.

12. A speech processing apparatus comprising:

a processor; and

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

determine an unvoicing/voicing parameter reflecting a characteristic of unvoiced/voiced speech in a current frame of a speech signal comprising a plurality of frames,

determine a smoothed unvoicing/voicing parameter to include information of the unvoicing/voicing parameter in a frame prior to the current frame of the speech signal,

compute a difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter, and

determine whether the current frame comprises unvoiced speech or voiced speech using the computed difference as a decision parameter.

13. The apparatus of claim 12, wherein the unvoicing/voicing parameter is a combined parameter reflecting a product of a periodicity parameter and a spectral tilt parameter.

14. The apparatus of claims 12 or 13, wherein when the difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter is greater than 0.1, determining the current frame of the speech signal to be an unvoiced/voiced signal, wherein when the difference between the unvoicing/voicing parameter and the smoothed unvoicing/voicing parameter is less than 0.05, determining the current frame of the speech signal to be not an unvoiced/voiced speech.

15. The apparatus of any of claims 12 to 14, wherein the unvoicing/voicing parameter is an unvoicing parameter reflecting a characteristic of unvoiced speech, and wherein the smoothed unvoicing/voicing parameter is a smoothed unvoicing parameter.

16. The apparatus of any of claims 12 to 14, wherein the unvoicing/voicing parameter is a voicing parameter reflecting a characteristic of voiced speech, and wherein the smoothed unvoicing/voicing parameter is a smoothed voicing parameter.

17. The apparatus of any one of claims 13 to 16, wherein the frame comprises a subframe.