

NOISE SIGNAL DETERMINING METHOD AND APPARATUS AND VOICE DENOISING METHOD AND APPARATUS

The present application claims priority to Chinese Patent Application No. 201510670697.8,
5 filed on October 13, 2015 and entitled "NOISE SIGNAL DETERMINING METHOD AND
APPARATUS AND VOICE DENOISING METHOD AND APPARATUS", which is
incorporated herein by reference in its entirety.

Technical Field

The present application relates to the field of voice denoising technologies, and in particular,
10 to a noise signal determining method and apparatus and a voice denoising method and apparatus.

Background Art

A voice denoising technology can improve voice quality by removing environment noises
from a voice signal. A power spectrum of a noise signal in a voice signal needs to be determined
first in the voice denoising process, and then the voice signal can be denoised according to the
15 determined power spectrum of the noise signal.

In the prior art, a power spectrum of a noise signal in a voice signal generally can be
determined in the following manner: analyzing first N frame signals in a voice signal segment on
the assumption that the first N frame signals are noise signals (i.e., including no human voice
signals), to obtain the power spectra of the noise signals in the voice signal.

20 In an actual application scenario, first N frame signals in a voice signal which are assumed to
be noise signals in the prior art are usually inconsistent with actual noise signals, and thus the
accuracy of obtained noise signal power spectra is affected.

Summary of the Invention

Objectives of embodiments of the present application are to provide a noise signal
25 determining method and apparatus and a voice denoising method and apparatus, to solve the
problem in the prior art that the accuracy of obtained noise signal power spectra is affected as
first N frame signals assumed to be noise signals are inconsistent with actual noise signals.

To solve the above technical problem, the noise signal determining method and apparatus and the voice denoising method and apparatus provided in the embodiments of the present application are implemented as follows:

A noise signal determining method, including:

- 5 performing Fourier transform on each frame signal in a to-be-analyzed voice signal segment to acquire a power spectrum of each frame signal in the voice signal segment;
- determining a variance of power values of each frame signal in the voice signal segment at various frequencies based on the power spectrum of the frame signal; and
- determining whether each frame signal in the voice signal segment is a noise signal based on
- 10 the variance.

A voice denoising method, including:

- determining a to-be-analyzed voice signal segment included in a to-be-processed voice;
- performing Fourier transform on each frame signal in the to-be-analyzed voice signal segment to acquire a power spectrum of each frame signal in the voice signal segment;
- 15 determining a variance of power values of each frame signal in the voice signal segment at various frequencies based on the power spectrum of the frame signal;
- determining whether each frame signal in the voice signal segment is a noise signal based on the variance to obtain several noise frames included in the voice signal segment; and
- determining an average power corresponding to the several noise frames included in the voice
- 20 signal segment, and denoising the to-be-processed voice based on the average power of the noise frames.

A noise signal determining apparatus, including:

- a power spectrum acquiring unit configured to perform Fourier transform on each frame signal in a to-be-analyzed voice signal segment to acquire a power spectrum of each frame signal
- 25 in the voice signal segment;
- a variance determining unit configured to determine a variance of power values of each frame signal in the voice signal segment at various frequencies based on the power spectrum of the

frame signal; and

a noise determining unit configured to determine whether each frame signal in the voice signal segment is a noise signal based on the variance.

A voice denoising apparatus, including:

5 a segment determining unit configured to determine a to-be-analyzed voice signal segment included in a to-be-processed voice;

a power spectrum acquiring unit configured to perform Fourier transform on each frame signal in the to-be-analyzed voice signal segment to acquire a power spectrum of each frame signal in the voice signal segment;

10 a variance determining unit configured to determine a variance of power values of each frame signal in the voice signal segment at various frequencies based on the power spectrum of the frame signal;

a noise determining unit configured to determine whether each frame signal in the voice signal segment is a noise signal based on the variance, and obtain several noise frames included

15 in the voice signal segment; and

a voice denoising unit configured to determine an average power corresponding to the several noise frames included in the voice signal segment, and denoise the to-be-processed voice based on the average power of the noise frames.

As can be seen from the above technical solutions provided in the embodiments of the
20 present application, by performing Fourier transform on a to-be-analyzed voice signal segment to acquire a power spectrum of each frame signal, determining a variance of power values of each frame signal in the to-be-analyzed voice signal segment at various frequencies, and finally determining whether the frame signal is a noise signal based on the variance, the noise signal determining method and apparatus as well as the voice denoising method and apparatus provided
25 in the embodiments of the present application can accurately obtain several noise frames included in the to-be-analyzed voice signal segment. The to-be-processed voice can be denoised based on an average power of the determined noise frames in the voice denoising process, and

thus the voice denoising effect is improved.

Brief Description of the Drawings

To describe the technical solutions in the embodiments of the present application or the prior art more clearly, the following briefly introduces the accompanying drawings used for describing the embodiments or the prior art. Apparently, the accompanying drawings described below are merely some embodiments recited in the present application, and those of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

FIG. 1 is a flowchart of a noise signal determining method according to an embodiment of the present application;

FIG. 2 is a flowchart of steps for determining whether a frame signal is a noise signal according to an embodiment of the present application;

FIG. 3 is a flowchart of steps for determining a variance of power values of a frame signal at various sampling points according to an embodiment of the present application;

FIG. 4 is a curve graph of variances of power values according to an embodiment of the present application;

FIG. 5 is a flowchart of a voice denoising method according to an embodiment of the present application;

FIG. 6 is a block diagram of a noise signal determining apparatus according to an embodiment of the present application;

FIG. 7 is a block diagram of a voice denoising apparatus according to an embodiment of the present application; and

FIG. 8 is a schematic structural diagram of a hardware implementation example of an apparatus provided in the present application.

Detailed Description

In order to make those skilled in the art better understand the technical solutions of the present application, the technical solutions in the embodiments of the present application will be described clearly and completely below with reference to the accompanying drawings in the

embodiments of the present application. It is apparent that the described embodiments are merely some, not all of the embodiments of the present application. Based on the embodiments of the present application, those of ordinary skill in the art can obtain other embodiments without creative efforts, which all fall within the protection scope of the present application.

5 FIG. 1 shows a flowchart of a noise signal determining method according to an embodiment of the present application. In order to determine a noise signal in a to-be-analyzed voice signal segment, the noise signal determining method of this embodiment includes the following steps:

 S101: Fourier transform is performed on each frame signal in the to-be-analyzed voice signal segment to acquire a power spectrum of each frame signal in the voice signal segment.

10 The to-be-analyzed voice signal segment can be captured from a to-be-processed voice based on a certain rule. The to-be-analyzed voice signal segment can be a "suspected noise frame segment" that possibly includes many noise frames based on preliminary determination. Preferably, before the step S101, the method further includes:

 determining a voice signal segment with an amplitude variation less than a preset threshold in
15 the to-be-processed voice as the to-be-analyzed voice signal segment based on an amplitude variation of a time-domain signal of the to-be-processed voice; or

 capturing first N frame voice signals in the to-be-processed voice as the to-be-analyzed voice signal segment.

 In the embodiment of the present application, in a time domain of a voice signal, a noise
20 signal is generally a voice signal segment having a small amplitude variation or having consistent amplitudes, while a voice signal segment including a human speech voice generally fluctuates greatly in amplitude variation. Based on such a rule, a preset threshold used for recognizing a "suspected noise frame segment" included in a to-be-processed voice (i.e., a to-be-denoised voice) may be set in advance. Therefore, a voice signal segment having an amplitude variation
25 less than the preset threshold in the to-be-processed voice can be determined as the to-be-analyzed voice signal segment.

 In the embodiment of the present application, framing can be performed on a voice signal

first. A frame signal refers to a single-frame voice signal, and one voice signal segment can include several frame signals. One frame signal can include several sampling points, e.g., 1024 sampling points. Two adjacent frame signals can overlap each other (for example, an overlap ratio can be 50%). In this embodiment, a short-time Fourier transform (STFT) can be performed
5 on a voice signal in a time domain to acquire a power spectrum (frequency domain) of the voice signal. The power spectrum can include multiple power values corresponding to different frequencies, e.g., 1024 power values.

In the embodiment of the present application, it can be generally assumed by default that a voice signal in a period of time (e.g., 1.5s) before a person speaks is a noise signal (an
10 environment noise) in a voice signal segment including a human voice. Therefore, it can be determined in the embodiment of the present application that the to-be-analyzed voice signal is first N frame signals in a voice signal segment. For example, the to-be-analyzed voice signal is a voice signal in the first 1.5s: $\{f_1', f_2', \dots, f_n'\}$, wherein $f_1', f_2', \dots, f_n'$ represent frame signals included in the voice signal respectively. The embodiment of the present application aims to determine
15 noise signals from the frame signals in the analyzed voice signal.

Multiple power values corresponding to each frame signal can be calculated based on the power spectrum of the to-be-analyzed voice signal: $\{f_1', f_2', \dots, f_n'\}$ obtained after the STFT. Assume that a power spectrum of a frame signal at a frequency is $a+bi$, wherein the real part a can represent the amplitude and the imaginary part b can represent the phase. Then a power value
20 of the frame signal at the frequency can be: a^2+b^2 . Power values of each frame signal at different frequencies can be obtained based on the above process. For example, if each of the frame signals $\{f_1', f_2', \dots, f_n'\}$ includes 1024 sampling points, 1024 power values of each frame signal at different frequencies can be obtained based on the power spectrum. For example, power values corresponding to the frame signal f_1' is $\{p^1_1, p^1_2, \dots, p^1_{1024}\}$, power values corresponding to the
25 frame signal f_2' is $\{p^2_1, p^2_2, \dots, p^2_{1024}\}$, ..., and power values corresponding to the frame signal f_n' is $\{p^n_1, p^n_2, \dots, p^n_{1024}\}$.

S102: A variance of power values of each frame signal in the voice signal segment at various

frequencies is determined based on the power spectrum of the frame signal.

Based on the power values of frame signals $\{f_1', f_2', \dots, f_n'\}$ at various frequencies, variances $\{Var(f_1'), Var(f_2'), \dots, Var(f_n')\}$ of the power values of the frame signals $\{f_1', f_2', \dots, f_n'\}$ can be calculated according to a variance calculation formula. For example, if each frame signal
5 includes 1024 sampling points, $Var(f_1')$ is a variance of $\{p^1_1, p^1_2, \dots, p^1_{1024}\}$, $Var(f_2')$ is a variance of $\{p^2_1, p^2_2, \dots, p^2_{1024}\}$, ..., and $Var(f_n')$ is a variance of $\{p^n_1, p^n_2, \dots, p^n_{1024}\}$.

S103: It is determined whether each frame signal in the voice signal segment is a noise signal based on the variance.

In the embodiment of the present application, energy (i.e., a power value) of a frame signal
10 including a speech segment generally varies with bands greatly, while energy of a frame signal without a speech segment (i.e., a noise signal) varies with bands slightly and is evenly distributed. Therefore, it can be determined whether each frame signal is a noise signal based on a variance of power values of the frame signal.

FIG. 2 shows a flowchart of steps for determining whether a frame signal is a noise signal
15 according to an embodiment of the present application. In the embodiment of the present application, the above step S103 can include the following steps:

S1031: It is determined whether the variance of the power values of the frame signal is greater than a first threshold T_1 .

S1032: If no, the frame signal is determined as a noise signal.

20 If a variance of power values of a frame signal exceeds the first threshold T_1 , it is indicated that a variation amplitude of energy (i.e., power values) of the frame signal with bands exceeds the first threshold T_1 . Therefore, it can be determined that the frame signal is not a noise signal. In contrast, if a variance of power values of a frame signal does not exceed the first threshold T_1 , it is indicated that a variation amplitude of energy (i.e., power values) of the frame signal with
25 bands does not exceed the first threshold T_1 . Therefore, it can be determined that the frame signal is a noise signal.

Based on the above process, noise frame signals $\{f_1', f_2', \dots, f_m'\}$ and non-noise frame signals $\{f$

$f'_{m+1}, f'_{m+2}, \dots, f'_n\}$ can be determined sequentially in the to-be-analyzed voice signals $\{f'_1, f'_2, \dots, f'_n\}$. Therefore, noise signals included in a voice signal segment can be determined, and voice denoising can be performed according to these noise signals $\{f'_1, f'_2, \dots, f'_m\}$.

Referring to FIG. 3, in the embodiment of the present application, the above step S102 can specifically include the following steps:

S1021: Power values of each of the frame signals $\{f'_1, f'_2, \dots, f'_n\}$ at various frequencies are at least classified into a first power value set corresponding to a first frequency interval and a second power value set corresponding to a second frequency interval according to frequency intervals to which frequencies corresponding to the power spectrum of the frame signal belong, the first frequency interval being lower than the second frequency interval.

In a specific embodiment, a variance of each frame signal can be acquired in the frequency domain through statistics. Non-noise signals are generally concentrated in low-mid frequency bands, while noise signals are generally distributed uniformly in all frequency bands. Therefore, a variance of power values of each frame signal at various frequencies can be acquired through statistics in at least two different frequency bands (i.e., the above frequency intervals).

For example, the first frequency interval can be 0~2000 Hz (low frequency band), and the second frequency interval can be 2000~4000 Hz (high frequency band). If each frame signal includes 1024 sampling points, 1024 power values corresponding to each frame signal are classified into a first power value set A corresponding to 0~2000 Hz and a second power value set B corresponding to 2000~4000 Hz according to the frequency intervals corresponding to the power values. Using the frame signal f'_1 as an example, 1024 corresponding power values are $\{p^1_1, p^1_2, \dots, p^1_{1024}\}$. According to the frequency intervals, it can be derived that power values included in the first power value set A are, for example, $\{p^1_1, p^1_2, \dots, p^1_{126}\}$, power values included in the first power set A are, for example, $\{p^1_{127}, p^1_{128}, \dots, p^1_{1024}\}$, and the rest can be deduced by analogy.

It should be noted that variances of signal power values can be acquired through statistics in more than two frequency bands in other embodiments of the present application.

S1022: A first variance of power values included in the first power value set is determined.

As described above, using the frame signal f_1' as an example, power values included in the first power value set A are, for example, $\{p_{127}^1, p_{128}^1, \dots, p_{1024}^1\}$. Therefore, a first variation $Var_{high}(f_1')$ of the power values $p_{127}^1 \sim p_{1024}^1$ can be calculated according to a variance formula.

5 S1021: A second variance of power values included in the second power value set is determined.

As described above, using the frame signal f_1' as an example, power values included in the second power value set B are, for example, $\{p_1^1, p_2^1, \dots, p_{126}^1\}$. Therefore, a second variation $Var_{low}(f_1')$ of the power values $p_1^1 \sim p_{126}^1$ can be calculated according to a variance formula.

10 FIG. 4 shows a schematic curve graph of variances according to an embodiment of the present application. In the graph, the horizontal axis indicates a frame number of a frame signal, and the vertical axis indicates the magnitude of a variance. A first variance curve shows the trend of a first variance of each frame signal, and the first variance curve shows the trend of a second variance of each frame signal. As can be seen from the graph that the variance fluctuates slightly
15 in the high frequency band 2000~4000 Hz, and the variance fluctuates greatly in the low frequency band 0~2000 Hz. This can prove that non-noise signals are mainly concentrated in the low frequency band.

As described above, in a preferred embodiment of the present application, the step S1031 can specifically include:

20 determining whether the first variance of the power values of the frame signal is greater than a first threshold T_1 ; and if yes, determining the frame signal as a noise signal. Using the frame signal f_1' as an example, it is determined whether the first variance $Var_{high}(f_1')$ is greater than the first threshold T_1 .

In the embodiment of the present application, the above step S103 can further specifically
25 include:

determining whether a difference between the first variance and the second variance is greater than a second threshold T_2 ; and

if no, determining the frame signal as a noise signal.

Using the frame signal f_1' as an example, a difference between the first variance and the second variance is $|Var_{high}(f_1') - Var_{low}(f_1')|$. If $|Var_{high}(f_1') - Var_{low}(f_1')| < T_2$, the frame signal f_1' is determined as a noise signal. Noise signals can be determined sequentially from the to-be-analyzed voice frame signals $\{f_1', f_2', \dots, f_n'\}$ according to this step.

In the embodiment of the present application, between the step S102 and the step S103, the method can further include:

ranking the frame signals in the to-be-analyzed voice signal segment according to magnitudes of the variances.

Then the determining whether each frame signal in the voice signal segment is a noise signal based on the variance includes:

determining whether each frame signal in the voice signal segment is a noise signal based on the variance of power values of each ranked frame signal at various frequencies.

As described above, variances $\{Var(f_1'), Var(f_2'), \dots, Var(f_n')\}$ of power values of the frame signals $\{f_1', f_2', \dots, f_n'\}$ can be determined in this embodiment. The frame signals can be ranked in ascending order of the variances of power values. A signal with a smaller variance is more likely a noise signal. Therefore, noise frame signals in the to-be-analyzed voice signal can be ranked to the front. In the embodiment of the present application, if variances are respectively acquired through statistics in the low frequency band (e.g., 0~2000 Hz) and the high frequency band (e.g., 2000~4000 Hz), power values of each of the frame signals $\{f_1', f_2', \dots, f_n'\}$ at various frequencies can be classified into a first power value set A corresponding to a first frequency interval (e.g., 0~2000 Hz) and a second power value set B corresponding to a second frequency interval (e.g., 2000~4000 Hz) according to the frequency intervals to which frequencies corresponding to the power spectrum of the frame signal belong. Then first variances $\{Var_{low}(f_1'), Var_{low}(f_2'), \dots, Var_{low}(f_n')\}$ of power values included in the first power value sets corresponding to the frame signals $\{f_1', f_2', \dots, f_n'\}$ can be determined respectively, and second variances $\{Var_{high}(f_1'), Var_{high}(f_2'), \dots, Var_{high}(f_n')\}$ of power values included in the second power value sets

corresponding to the frame signals $\{f_1', f_2', \dots, f_n'\}$ can be determined respectively. In the step S104 above, based on the variance statistics at high frequencies and low frequencies, noise signals included in the to-be-analyzed voice signals (which can be voice signals ranked according to magnitudes of variances) can be determined in the following manner:

$$5 \quad Var_{low}(f_i') > T_1 \quad (1);$$

$$|Var_{high}(f_i') - Var_{low}(f_i')| > T_2 \quad (2);$$

$$Var_{high}(f_{i+1}') - Var_{high}(f_{i-1}') > T_3 \quad (3);$$

$$Var_{high}(f_{i+1}') - Var_{low}(f_{i-1}') > T_4 \quad (4);$$

$i \in (1, n)$. It can be determined based on formula (1) whether a first variance of power values of each frame signal f_i' is greater than a first threshold T_1 . If no, the frame signal f_i' is determined as a noise frame signal. A set of determined noise frame signals is determined as a noise signal.

It can be determined based on formula (2) whether a second variance of power values of each frame signal f_i' is greater than a second threshold T_2 . If no, the frame signal f_i' is determined as a noise frame signal. A set of determined noise frame signals is determined as a noise signal.

15 It can be determined based on formula (3) whether a difference $Var_{high}(f_{i+1}') - Var_{high}(f_{i-1}')$ between a second variance $Var_{high}(f_{i-1}')$ of power values of a frame signal f_{i-1}' prior to a frame signal f_i' and a second variance $Var_{high}(f_{i+1}')$ of power values of a frame signal f_{i+1}' next to the frame signal f_i' is greater than a third threshold T_3 . If no, the frame signal f_i' is determined as a noise frame signal. A set of determined noise frame signals is determined as a noise signal.

20 It can be determined based on formula (4) whether a difference $Var_{low}(f_{i+1}') - Var_{low}(f_{i-1}')$ between a first variance $Var_{low}(f_{i-1}')$ of power values of a frame signal f_{i-1}' prior to a frame signal f_i' and a first variance $Var_{low}(f_{i+1}')$ of power values of a frame signal f_{i+1}' next to the frame signal f_i' is greater than a fourth threshold T_4 . If no, the frame signal f_i' is determined as a noise frame signal. A set of determined noise frame signals is determined as a noise signal.

25 In the embodiment of the present application, noise frames included in the to-be-analyzed voice signal can be recognized by using the above formulas (1) to (4). That is, any frame signal f_i' meeting any one of the above formulas (1) to (4) can be determined as a non-noise signal (a

noise end frame). In other words, any frame signal f_i' meeting none of the above formulas (1) to (4) can be determined as a noise signal. A noise end frame f_m' can be determined based on the above process, and then the noise frames include: $\{f_1', f_2', \dots, f_{m-1}'\}$.

It should be noted that, in other embodiments of the present application, the noise end frame can be determined based on some of the formulas (1) to (4), such as the formulas (1) and (2), or the formulas (2) and (3). Moreover, formulas for determining the noise end frame in the embodiment of the present application are not limited to the formulas listed above. The thresholds T_1 , T_2 , T_3 and T_4 are all obtained from statistics on a large quantity of testing samples.

FIG. 5 is a flowchart of a voice denoising method according to an embodiment of the present application, including the following steps:

S201: A to-be-analyzed voice signal segment included in a to-be-processed voice is determined.

S202: Fourier transform is performed on each frame signal in the to-be-analyzed voice signal segment to acquire a power spectrum of each frame signal in the voice signal segment.

S203: A variance of power values of each frame signal in the voice signal segment at various frequencies is determined based on the power spectrum of the frame signal.

S204: It is determined whether each frame signal in the voice signal segment is a noise signal based on the variance, and several noise frames included in the voice signal segment are obtained.

S205: An average power corresponding to the several noise frames included in the voice signal segment is determined, and the to-be-processed voice is denoised based on the average power of the noise frames.

In the embodiment of the present application, after noise frames $\{f_1', f_2', \dots, f_{m-1}'\}$ included in a to-be-analyzed voice segment are acquired according to the above method, frame numbers of original signals (before ranking) corresponding to the noise frames respectively can be determined, and an average power of these frame signals can be obtained through statistics to obtain a power spectrum estimation value P_{noise} of the noise signal. The voice can be denoised

after the power spectrum estimation value P_{noise} of the noise signal is obtained. The denoising method is well known to those of ordinary skill in the art and will not be described specifically here.

Definitely, in other feasible embodiments of the present application, the step of ranking the
5 frame signals according to the variances may be omitted, and noise frames can be determined directly based on variances of the original signals. In addition, after multiple frames of noise signal are determined in the present application, the power spectrum estimation value P_{noise} is generally calculated by using some of the frames, to avoid over-estimation. For example, first 30 frames can be captured to calculate the power spectrum estimation value P_{noise} if the determined
10 noise signal includes 50 frames. As such, the accuracy of the power spectrum estimation value can be improved.

An embodiment of the present application further provides a noise signal determining apparatus corresponding to the above process implementation. The apparatus can be implemented through software, and can also be implemented through hardware or a combination
15 of software and hardware. By using a software implementation manner as an example, an apparatus in a logic sense can be formed by reading a corresponding computer program through a Central Process Unit (CPU) of a server into a memory and running the computer program. Refer to FIG. 8 for a hardware structure of the apparatus.

FIG. 6 is a block diagram of a noise signal determining apparatus according to an
20 embodiment of the present application. In this embodiment, functions of units in the apparatus can correspond to functions of the steps in the above noise signal determining method. Refer to the above method embodiment for details. The noise signal determining apparatus 100 includes:

a power spectrum acquiring unit 101 configured to perform Fourier transform on each frame signal in a to-be-analyzed voice signal segment to acquire a power spectrum of each frame signal
25 in the voice signal segment;

a variance determining unit 102 configured to determine a variance of power values of each frame signal in the voice signal segment at various frequencies based on the power spectrum of

the frame signal; and

a noise determining unit 103 configured to determine whether each frame signal in the voice signal segment is a noise signal based on the variance.

Preferably, the apparatus further includes: a segment acquiring unit configured to:

- 5 determine a voice signal segment with an amplitude variation less than a preset threshold in a to-be-processed voice as the to-be-analyzed voice signal segment based on an amplitude variation of a time-domain signal of the to-be-processed voice; or

capture first N frame voice signals in a to-be-processed voice as the to-be-analyzed voice signal segment.

- 10 Preferably, the noise determining unit 103 is configured to:

determine whether the variance corresponding to each frame signal in the voice signal segment is greater than a first threshold; and

if no, determine the frame signal as a noise signal.

Preferably, the variance determining unit 102 is configured to:

- 15 at least classify power values of the frame signal at various frequencies into a first power value set corresponding to a first frequency interval according to frequency intervals to which frequencies corresponding to the power spectrum belong; and

determine a first variance of power values included in the first power value set.

Then the noise determining unit 103 is configured to:

- 20 determine whether the first variance is greater than the first threshold; and

if no, determine the frame signal as a noise signal.

Preferably, the variance determining unit 102 is specifically configured to:

- at least classify power values of each frame signal at various frequencies into a first power value set corresponding to a first frequency interval and a second power value set corresponding
25 to a second frequency interval according to frequency intervals to which frequencies corresponding to the power values of the frame signal belong, wherein the first frequency interval is lower than the second frequency interval;

determine a first variance of power values included in the first power value set; and
determine a second variance of power values included in the second power value set.

Then the noise determining unit 103 is configured to:

determine whether a difference between the first variance and the second variance which
5 correspond to each frame signal is greater than a second threshold; and
if no, determine the frame signal as a noise signal.

An embodiment of the present application further provides a voice denoising apparatus
corresponding to the above process implementation. The apparatus can be implemented through
software, and can also be implemented through hardware or a combination of software and
10 hardware. By using a software implementation manner as an example, an apparatus in a logic
sense can be formed by reading a corresponding computer program through a Central Process
Unit (CPU) of a server into a memory and running the computer program. Refer to FIG. 8 for a
hardware structure of the apparatus.

FIG. 7 is a block diagram of a voice denoising apparatus according to an embodiment of the
15 present application. In this embodiment, functions of units in the apparatus can correspond to
functions of the steps in the above voice denoising method. Refer to the above method
embodiment for details. In this embodiment, the voice denoising apparatus 200 includes:

a segment determining unit 201 configured to determine a to-be-analyzed voice signal
segment included in a to-be-processed voice;

20 a power spectrum acquiring unit 202 configured to perform Fourier transform on each frame
signal in the to-be-analyzed voice signal segment to acquire a power spectrum of each frame
signal in the voice signal segment;

a variance determining unit 203 configured to determine a variance of power values of each
frame signal in the voice signal segment at various frequencies based on the power spectrum of
25 the frame signal;

a noise determining unit 205 configured to determine whether each frame signal in the voice
signal segment is a noise signal based on the variance, and obtain several noise frames included

in the voice signal segment; and

a voice denoising unit 10 configured to determine an average power corresponding to the several noise frames included in the voice signal segment, and denoise the to-be-processed voice based on the average power of the noise frames.

5 Preferably, the apparatus further includes: a ranking unit 204 configured to:

rank the frame signals in the to-be-analyzed voice signal segment according to magnitudes of the variances.

Then the noise determining unit 205 is specifically configured to:

10 determine whether each frame signal in the voice signal segment is a noise signal based on the variance of power values of each ranked frame signal at various frequencies.

By performing Fourier transform on a to-be-analyzed voice signal segment to acquire a power spectrum of each frame signal, determining a variance of power values of each frame signal in the to-be-analyzed voice signal segment at various frequencies, and finally determining whether the frame signal is a noise signal based on the variance, the noise signal determining
15 method and apparatus as well as the voice denoising method and apparatus provided in the embodiments of the present application can accurately determine several noise frames included in the to-be-analyzed voice signal segment. The to-be-processed voice can be denoised based on an average power of the determined several noise frames in the voice denoising process, and thus the voice denoising effect is improved.

20 For ease of description, the apparatus is divided into various units in terms of functions for respective descriptions. Definitely, when the present application is implemented, functions of the units may be implemented in the same software and/or hardware component or multiple software and/or hardware components.

Those skilled in the art should understand that the embodiments of the present invention may
25 be provided as a method, a system, or a computer program product. Therefore, the present invention may be implemented as a complete hardware embodiment, a complete software embodiment, or an embodiment combining software and hardware. Moreover, the present

invention may be in the form of a computer program product implemented on one or more computer usable storage media (including, but not limited to, a magnetic disk memory, a CD-ROM, an optical memory, and the like) including computer usable program codes.

5 The present invention is described with reference to flowcharts and/or block diagrams according to the method, the device (system) and the computer program product according to the embodiments of the present invention. It should be understood that a computer program instruction may be used to implement each process and/or block and a combination of processes and/or blocks in the flowcharts and/or block diagrams. The computer program instructions may be provided to a general-purpose computer, a special-purpose computer, an embedded processor
10 or other programmable data processing device to generate a machine, such that the computer or a processor of other programmable data processing device executes an instruction to generate an apparatus configured to implement functions designated in one or more processes in a flowchart and/or one or more blocks in a block diagram.

The computer program instructions may also be stored in a computer readable storage that
15 can guide a computer or other programmable data processing device to work in a specific manner, such that the instruction stored in the computer readable storage generates a manufacture including an instruction apparatus which implements functions designated by one or more processes in a flowchart and/or one or more blocks in a block diagram.

The computer program instructions may also be loaded in a computer or other programmable
20 data processing device, such that a series of operation steps are executed on the computer or other programmable device to generate a computer implemented processing. Therefore, the instruction executed in the computer or other programmable device provides steps for implementing functions designated in one or more processes in a flowchart and/or one or more blocks in a block diagram.

25 It should be further noted that the term "include", "comprise" or other variations thereof are intended to cover non-exclusive including, so that a process, method, commodity or device including a series of elements not only includes the elements, but also includes other elements

not clearly listed, or further includes inherent elements of the process, method, commodity or device. In a case without any more limitations, an element defined by "including a/an..." does not exclude that the process, method, commodity or device including the element further has other identical elements.

5 Those skilled in the art should understand that the embodiments of the present application may be provided as a method, a system, or a computer program product. Therefore, the present application may be implemented in the form of a complete hardware embodiment, a complete software embodiment, or an embodiment combining software and hardware. Moreover, the present application may be in the form of a computer program product implemented on one or
10 more computer usable storage media (including, but not limited to, a magnetic disk memory, a CD-ROM, an optical memory, and the like) including computer usable program codes.

 The present application may be described in a common context of a computer executable instruction executed by a computer, for example, a program module. Generally, the program module includes a routine, a program, an object, an assembly, a data structure, and the like used
15 for executing a specific task or implementing a specific abstract data type. The present application may also be implemented in distributed computing environments, in which a task is executed by using remote processing devices connected through a communications network. In the distributed computer environments, the program module may be located in local and remote computer storage media including a storage device.

20 The embodiments in the specification are described progressively, identical or similar parts of the embodiments may be obtained with reference to each other, and each embodiment emphasizes a part different from other embodiments. Especially, the system embodiment is basically similar to the method embodiment, so it is described simply. For related parts, refer to the descriptions of the parts in the method embodiment.

25 The above descriptions are merely embodiments of the present application, and are not intended to limit the present application. Various modifications and variations of the present application are possible to those skilled in the art. Any modification, equivalent replacement,

improvement and the like made within the spirit and principle of the present application should all fall within the scope of claims of the present application.

CLAIMS

1. A method for determining noise signals in a voice signal segment, the method comprising:

performing, on each frame signal in the voice signal segment, a frequency domain
5 transform to obtain a power spectrum of the frame signal, wherein the power spectrum comprises multiple power values corresponding to different frequencies;

classifying power values of each frame signal at various frequencies into a first power value set corresponding to a first frequency interval and a second power value set corresponding to a second frequency interval according to frequency intervals to which frequencies
10 corresponding to the power values of the frame signal belong, wherein the first frequency interval is lower than the second frequency interval;

determining a first variance of power values included in the first power value set;

determining a second variance of power values included in the second power value set;

determining whether a difference between the first variance and the second variance
15 which correspond to each frame signal is greater than a second threshold; and

in response to determining that the difference between the first variance and the second variance is not greater than a second threshold, identifying the frame signal as a noise signal.

2. The method of claim 1, further comprising:

20 extracting the voice signal segment from an input voice;

obtaining multiple noise frame signals included in the voice signal segment based on determining whether each frame signal in the voice signal segment is a noise signal;

determining an average power corresponding to the multiple noise frame signals included in the voice signal segment; and

25 performing noise reduction on the input voice based on the average power of the noise frame signals.

3. The method of claim 2, wherein the step of extracting the voice signal segment from an input voice comprises:

30 identifying a voice signal segment with an amplitude variation less than a preset threshold in the input voice as the voice signal segment based on an amplitude variation of a time-domain signal of the input voice; or

capturing a first N frame voice signals in the input voice and using the captured first N frame voice signals as the voice signal segment.

35

4. The method of claim 2, wherein determining the average power corresponding to the multiple noise frame signals included in the voice signal segment comprises calculating the average power using a number of noise frame signals that is less than the total number of noise frame signals to avoid over estimation of the average power.
- 5
5. The method of claim 1, further comprising:
identifying a voice signal segment with an amplitude variation less than a preset threshold in a input voice as the voice signal segment based on an amplitude variation of a time-domain signal of the input voice; or
10 capturing a first N frame voice signals in a input voice and using the captured first N frame voice signals as the voice signal segment.
6. The method of claim 1 or claim 2, wherein identifying the frame signal as the noise signal comprises:
15 determining whether the first variance corresponding to each frame signal in the voice signal segment is greater than a first threshold; and
in response to determining that the first variance corresponding to a frame signal is not greater than the first threshold, identifying the frame signal as a noise signal.
- 20 7. The method of claim 6, wherein determining the first variance of power values comprises:
classifying power values of the frame signal at various frequencies into a first power value set corresponding to a first frequency interval according to frequency intervals to which frequencies corresponding to the power spectrum belong; and
determining the first variance of power values comprised in the first power value set;
25 then the step of determining whether the first variance is greater than a first threshold comprises determining whether the first variance is greater than the first threshold.
8. The method of claim 1 or claim 2, wherein determining the first variance of power values comprises:
30 classifying power values of each frame signal at various frequencies into a first power value set corresponding to a first frequency interval and a second power value set corresponding to a second frequency interval according to frequency intervals to which frequencies corresponding to the power values of the frame signal belong, wherein the first frequency interval is lower than the second frequency interval;
35 determining a first variance of power values included in the first power value set; and

determining a second variance of power values included in the second power value set;
then the step of determining whether each frame signal in the voice signal segment is a noise signal based on the variances comprises:

5 determining whether a difference between the first variance and the second variance which correspond to each frame signal is greater than a second threshold; and

in response to determining that the difference between the first variance and the second variance is not greater than a second threshold, identifying the frame signal as a noise signal.

9. The method of claim 1 or claim 2, further comprising:

10 ranking the frame signals in the voice signal segment according to magnitudes of the variances,

wherein determining whether each frame signal in the voice signal segment is a noise signal based on the variances comprises determining whether each frame signal in the voice signal segment is a noise signal based on the variance of power values of each ranked frame
15 signal at various frequencies.

10. An apparatus for determining noise signals in a voice signal segment, the apparatus comprising multiple units configured to perform the method of any one of claims 1 to 9.