CLAIMS

1. A computer-implemented method, comprising:

obtaining, by a user terminal, an audio signal;

determining a ratio of a sampling rate of the audio signal to a frequency of a predetermined voice signal;

dividing, by the user terminal, the audio signal into a maximum quantity of short-time energy frames containing a plurality of samples based on the ratio;

determining, by the user terminal, energy of each short-time energy frame; and

determining, by the user terminal, whether the audio signal includes a voice signal based on the energy of each short-time energy frame.

2. The computer-implemented method of claim 1, wherein the audio signal is collected at the sampling rate and is in a pulse code modulation (PCM) format.

3. The computer-implemented method of claim 1, wherein the obtained audio signal is in a non-PCM format, and further comprising:

prior to dividing the audio signal:

converting the audio signal into a pulse code modulation (PCM) format; and

identifying the sampling rate of the audio signal.

4. The computer-implemented method of claim 1, wherein the energy of each short-time energy frame is a sum of energy associated with each sampling point in each short-time energy frame, and wherein the energy associated with each sampling point is determined based on an amplitude of the audio signal that corresponds to the sampling point in the short-time energy frame.

5. The computer-implemented method of claim 1, wherein determining whether the audio signal includes a voice signal comprises:

determining a plurality of high-energy frames, wherein each high-energy frame of the plurality of high-energy frames is a short-time energy frame where energy is greater than a predetermined threshold;

determining a high-energy frame ratio that is represented by a ratio of a quantity of the plurality of high-energy frames to a quantity of the short-time energy frames included in the audio signal;

determining whether the high-energy frame ratio is greater than a predetermined value;

if it is determined that the high-energy frame ratio is greater than the predetermined value:

determining that the audio signal includes a voice signal; or

if it is determined that the high-energy frame ratio is not greater than the predetermined value:

determining that the audio signal does not include a voice signal.

6. The computer-implemented method of claim 5, wherein it is determined that the high-energy frame ratio is greater than the predetermined value, further comprising:

determining, from the short-time energy frames included in the audio signal, whether there exist a predetermined number of consecutive short-time energy frames, wherein each of the predetermined number of consecutive short-time energy frame has energy that is greater than the predetermined threshold;

if YES, determining that the audio signal includes a voice signal; or

otherwise, determining that the audio signal does not include a voice signal.

7. A non-transitory, computer-readable medium storing one or more instructions executable by a computer system to perform operations comprising:

obtaining, by a user terminal, an audio signal;

determining a ratio of a sampling rate of the audio signal to a frequency of a predetermined voice signal;

dividing, by the user terminal, the audio signal into a maximum quantity of short-time energy frames containing a plurality of samples based on the ratio;

determining, by the user terminal, energy of each short-time energy frame; and

determining, by the user terminal, whether the audio signal includes a voice signal based on the energy of each short-time energy frame.

8. The non-transitory, computer-readable medium of claim 7, wherein the audio signal is collected at the sampling rate and is in a pulse code modulation (PCM) format.

9. The non-transitory, computer-readable medium of claim 7, wherein the obtained audio signal is in a non-PCM format, and further comprising:

prior to dividing the audio signal:

converting the audio signal into a pulse code modulation (PCM) format; and

identifying the sampling rate of the audio signal.

10. The non-transitory, computer-readable medium of claim 7, wherein the energy of each short-time energy frame is a sum of energy associated with each sampling point in each short-time energy frame, and wherein the energy associated with each sampling point is determined based on an amplitude of the audio signal that corresponds to the sampling point in the short-time energy frame.

11. The non-transitory, computer-readable medium of claim 7, wherein determining whether the audio signal includes a voice signal comprises:

determining a plurality of high-energy frames, wherein each high-energy frame of the plurality of high-energy frames is a short-time energy frame where energy is greater than a predetermined threshold;

determining a high-energy frame ratio that is represented by a ratio of a quantity of the plurality of high-energy frames to a quantity of the short-time energy frames included in the audio signal;

determining whether the high-energy frame ratio is greater than a predetermined value;

if it is determined that the high-energy frame ratio is greater than the predetermined value:

determining that the audio signal includes a voice signal; or

if it is determined that the high-energy frame ratio is not greater than the predetermined value:

determining that the audio signal does not include a voice signal.

12. The non-transitory, computer-readable medium of claim 11, wherein it is determined that the high-energy frame ratio is greater than the predetermined value, further comprising:

determining, from the short-time energy frames included in the audio signal, whether there exist a predetermined number of consecutive short-time energy frames, wherein each of the predetermined number of consecutive short-time energy frame has energy that is greater than the predetermined threshold;

if YES, determining that the audio signal includes a voice signal; or

otherwise, determining that the audio signal does not include a voice signal.

13. A computer-implemented system, comprising:

one or more computers; and

one or more computer memory devices interoperably coupled with the one or more computers and having tangible, non-transitory, machine-readable media storing one or more instructions that, when executed by the one or more computers, perform one or more operations comprising:

obtaining, by a user terminal, an audio signal;

determining a ratio of a sampling rate of the audio signal to a frequency of a predetermined voice signal;

dividing, by the user terminal, the audio signal into a maximum quantity of short-time energy frames containing a plurality of samples based on the ratio;

determining, by the user terminal, energy of each short-time energy frame; and

determining, by the user terminal, whether the audio signal includes a voice signal based on the energy of each short-time energy frame.

14. The computer-implemented system of claim 13, wherein the audio signal is collected at the sampling rate and is in a pulse code modulation (PCM) format.

15. The computer-implemented system of claim 13, wherein the obtained audio signal is in a non-PCM format, and further comprising:

prior to dividing the audio signal:

converting the audio signal into a pulse code modulation (PCM) format; and

identifying the sampling rate of the audio signal.

16. The computer-implemented system of claim 13, wherein the energy of each short-time energy frame is a sum of energy associated with each sampling point in each short-time energy frame, and wherein the energy associated with each sampling point is determined based on an amplitude of the audio signal that corresponds to the sampling point in the short-time energy frame.

17. The computer-implemented system of claim 13, wherein determining whether the audio signal includes a voice signal comprises:

determining a plurality of high-energy frames, wherein each high-energy frame of the plurality of high-energy frames is a short-time energy frame where energy is greater than a predetermined threshold;

determining a high-energy frame ratio that is represented by a ratio of a quantity of the plurality of high-energy frames to a quantity of the short-time energy frames included in the audio signal;

determining whether the high-energy frame ratio is greater than a predetermined value;

if it is determined that the high-energy frame ratio is greater than the predetermined value:

determining that the audio signal includes a voice signal; or

if it is determined that the high-energy frame ratio is not greater than the predetermined value:

determining that the audio signal does not include a voice signal.