CLAIMS

1. A computer-implemented method for voice signal detection, the computer-implemented method comprising:

obtaining an audio signal;

dividing the audio signal into a plurality of short-time energy frames based on a frequency of a predetermined voice signal comprising a human voice, the frequency being set to a minimum human voice frequency;

determining energy of each short-time energy frame;

determining, based on the energy of each short-time energy frame, whether the audio signal comprises a voice signal; and

in response to determining that the audio signal comprises a voice signal providing an output comprising the voice signal.

2. The computer-implemented method of claim 1, wherein obtaining the audio signal comprises:

obtaining a current audio signal;

clipping a sub-signal with a predetermined time period from a last obtained audio signal; and

concatenating the current audio signal and the clipped sub-signal, to serve as the obtained audio signal.

3. The computer-implemented method of claim 1, wherein dividing the audio signal into the plurality of short-time energy frames comprises:

determining a period of the predetermined voice signal based on the frequency of the predetermined voice signal; and

dividing, based on the determined period, the audio signal into a plurality of short-time energy frames whose corresponding duration is the period.

4. The computer-implemented method of claim 1, wherein detecting, based on the energy of each short-time energy frame, whether the audio signal comprises the voice signal comprises:

determining a ratio of a quantity of short-time energy frames whose energy is greater than a predetermined threshold to a total quantity of all short-time energy frames; and

determining whether the ratio is greater than a predetermined ratio.

5. The computer-implemented method of claim 4, wherein:

in response to determining that the ratio is greater than the predetermined ratio, determining that the audio signal comprises a voice signal.

6. The computer-implemented method of claim 4, wherein:

in response to determining that the ratio is not greater than the predetermined ratio, determining that the audio signal does not comprise a voice signal.

7. The computer-implemented method of claim 4, wherein

in response to determining that the ratio is not greater than the predetermined ratio, determining that the audio signal does not comprise a voice signal.

8. The computer-implemented method of claim 4, wherein:

in response to determining that the ratio is greater than the predetermined ratio, determining whether the at least N consecutive short-time energy frames in the short-time energy frames comprise an energy that is greater than the predetermined threshold; and

in response to determining that the at least N consecutive short-time energy frames in the short-time energy frames whose energy is greater than the predetermined threshold, determining that the audio signal comprises a voice signal.

9. The computer-implemented method of claim 4, wherein:

in response to determining that the ratio is greater than the predetermined ratio, determining whether the at least N consecutive short-time energy frames in the short-time energy frames comprise an energy that is greater than the predetermined threshold; and

in response to determining that there are not at least N consecutive short-time energy frames in the short-time energy frames whose energy is greater than the predetermined threshold, determining that the audio signal does not comprise a voice signal.

10. The computer-implemented method of claim 1, further comprising:

determining a start point and an end point of the voice signal;

adding an audio stop signal to the voice signal after the end point, to indicate that the voice signal is complete; and

transmitting the voice signal for subsequent processing.

11. An apparatus for voice signal detection, the apparatus comprising a plurality of modules configured to perform the computer-implemented method of any one of claims 1 to 10.