CLAIMS

1. An audio signal classification method, comprising:

determining, according to voice activity of a current audio frame, whether to obtain a frequency spectrum fluctuation of the current audio frame and store the frequency spectrum fluctuation in a frequency spectrum fluctuation memory, wherein the frequency spectrum fluctuation denotes an energy fluctuation of a frequency spectrum of an audio signal;

updating, according to whether the audio frame is percussive music or activity of a historical audio frame, frequency spectrum fluctuations stored in the frequency spectrum fluctuation memory; and

classifying the current audio frame as a speech frame or a music frame according to statistics of a part or all of effective data of the updated frequency spectrum fluctuations stored in the frequency spectrum fluctuation memory.

2. The method according to claim 1, wherein the determining, according to voice activity of a current audio frame, whether to obtain a frequency spectrum fluctuation of the current audio frame and store the frequency spectrum fluctuation in a frequency spectrum fluctuation memory comprises:

if the current audio frame is an active frame, storing the frequency spectrum fluctuation of the current audio frame in the frequency spectrum fluctuation memory.

3. The method according to claim 1, wherein the determining, according to voice activity of a current audio frame, whether to obtain a frequency spectrum fluctuation of the current audio frame and store the frequency spectrum fluctuation in a frequency spectrum fluctuation memory comprises:

if the current audio frame is an active frame, and the current audio frame does not belong to an energy attack, storing the frequency spectrum fluctuation of the current audio frame in the frequency spectrum fluctuation memory.

4. The method according to claim 1, wherein the determining, according to voice activity of a current audio frame, whether to obtain a frequency spectrum fluctuation of the current audio frame and store the frequency spectrum fluctuation in a frequency spectrum fluctuation memory comprises:

if the current audio frame is an active frame, and none of multiple consecutive frames comprising the current audio frame and a historical frame of the current audio frame belongs to an energy attack, storing the frequency spectrum fluctuation of the current audio frame in the

frequency spectrum fluctuation memory.

5. The method according to any one of claims 1 to 4, wherein the updating, according to activity of a historical audio frame, frequency spectrum fluctuations stored in the frequency spectrum fluctuation memory comprises:

if it is determined that the frequency spectrum fluctuation of the current audio frame is stored in the frequency spectrum fluctuation memory, and a previous audio frame is an inactive frame, modifying data of other frequency spectrum fluctuations stored in the frequency spectrum fluctuation memory except the frequency spectrum fluctuation of the current audio frame into ineffective data; or

if it is determined that the frequency spectrum fluctuation of the current audio frame is stored in the frequency spectrum fluctuation memory, and three consecutive historical frames before the current audio frame are not all active frames, modifying the frequency spectrum fluctuation of the current audio frame into a first value; or

if it is determined that the frequency spectrum fluctuation of the current audio frame is stored in the frequency spectrum fluctuation memory, and a historical classification result is a music signal and the frequency spectrum fluctuation of the current audio frame is greater than a second value, modifying the frequency spectrum fluctuation of the current audio frame into the second value.

6. The method according to any one of claims 1 to 4, wherein the updating, according to whether the current audio frame is percussive music, frequency spectrum fluctuations stored in the frequency spectrum fluctuation memory comprises:

if the current audio frame belongs to percussive music, modifying values of the frequency spectrum fluctuations stored in the frequency spectrum fluctuation memory.

7. The method according to any one of claims 1 to 6, wherein the classifying the current audio frame as a speech frame or a music frame according to statistics of a part or all of effective data of the updated frequency spectrum fluctuations stored in the frequency spectrum fluctuation memory comprises:

obtaining an average value of a part or all of the effective data of the frequency spectrum fluctuations stored in the updated frequency spectrum fluctuation memory; and

when the obtained average value of the effective data of the updated frequency spectrum fluctuations satisfies a music classification condition, classifying the current audio frame as a music frame.

8. The method according to any one of claims 1 to 6, further comprising:

obtaining a frequency spectrum high-frequency-band peakiness, a frequency spectrum correlation degree, and a linear prediction residual energy tilt of the current audio frame, wherein

the frequency spectrum high-frequency-band peakiness denotes a peakiness or an energy acutance, on a high frequency band, of a frequency spectrum of the current audio frame; the frequency spectrum correlation degree denotes stability, between adjacent frames, of a signal harmonic structure of the current audio frame; and the linear prediction residual energy tilt denotes an extent to which linear prediction residual energy of the audio signal changes as a linear prediction order increases; and

determining, according to the voice activity of the current audio frame, whether to store the frequency spectrum high-frequency-band peakiness, the frequency spectrum correlation degree, and the linear prediction residual energy tilt in a memory,

wherein the classifying the audio frame according to statistics of a part or all of effective data of the frequency spectrum fluctuations stored in the frequency spectrum fluctuation memory comprises:

obtaining an average value of the effective data of the stored frequency spectrum fluctuations, an average value of effective data of stored frequency spectrum high-frequency-band peakiness, an average value of effective data of stored frequency spectrum correlation degrees, and a variance of effective data of stored linear prediction residual energy tilts separately; and

when a classifying condition comprising one of the following conditions is satisfied, classifying the current audio frame as a music frame; otherwise classifying the current audio frame as a speech frame: the average value of the effective data of the frequency spectrum fluctuations is less than a first threshold; or the average value of the effective data of the frequency spectrum high-frequency-band peakiness is greater than a second threshold; or the average value of the effective data of the frequency spectrum correlation degrees is greater than a third threshold; or the variance of the effective data of the linear prediction residual energy tilts is less than a fourth threshold.

9. An audio signal classification apparatus, wherein the apparatus is configured to classify an input audio signal, and comprises:

a storage determining unit, configured to determine, according to voice activity of a current audio frame, whether to obtain and store a frequency spectrum fluctuation of the current audio frame, wherein the frequency spectrum fluctuation denotes an energy fluctuation of a frequency spectrum of an audio signal;

a memory, configured to store the frequency spectrum fluctuation when the storage determining unit outputs a result that the frequency spectrum fluctuation needs to be stored;

an updating unit, configured to update, according to whether the audio frame is percussive music or activity of a historical audio frame, frequency spectrum fluctuations stored in the memory;

and

a classification unit, configured to classify the current audio frame as a speech frame or a music frame according to statistics of a part or all of effective data of the updated frequency spectrum fluctuations stored in the memory.

10. The apparatus according to claim 9, wherein the storage determining unit is specifically configured to: when it is determined that the current audio frame is an active frame, output a result that the frequency spectrum fluctuation of the current audio frame needs to be stored.

11. The apparatus according to claim 9, wherein the storage determining unit is specifically configured to: when it is determined that the current audio frame is an active frame, and the current audio frame does not belong to an energy attack, output a result that the frequency spectrum fluctuation of the current audio frame needs to be stored.

12. The apparatus according to claim 9, wherein the storage determining unit is specifically configured to: when it is determined that the current audio frame is an active frame, and none of multiple consecutive frames comprising the current audio frame and a historical frame of the current audio frame belongs to an energy attack, output a result that the frequency spectrum fluctuation of the current audio frame needs to be stored.

13. The apparatus according to any one of claims 9 to 12, wherein the updating unit is specifically configured to: if the current audio frame is an active frame, and a previous audio frame is an inactive frame, modify data of other frequency spectrum fluctuations stored in the memory except the frequency spectrum fluctuation of the current audio frame into ineffective data; or

if the current audio frame is an active frame, and three consecutive frames before the current audio frame are not all active frames, modify the frequency spectrum fluctuation of the current audio frame into a first value; or

if the current audio frame is an active frame, and a historical classification result is a music signal and the frequency spectrum fluctuation of the current audio frame is greater than a second value, modify the frequency spectrum fluctuation of the current audio frame into the second value.

14. The apparatus according to any one of claims 9 to 12, wherein the updating unit is specifically configured to: if the current audio frame belongs to percussive music, modify values of the frequency spectrum fluctuations stored in the frequency spectrum fluctuation memory.

15. The apparatus according to any one of claims 9 to 14, wherein the classification unit comprises:

a calculating unit, configured to obtain an average value of a part or all of the effective data of the updated frequency spectrum fluctuations stored in the memory; and

a determining unit, configured to compare the average value of the effective data of the

frequency spectrum fluctuations with a music classification condition; and when the average value of the effective data of the updated frequency spectrum fluctuations satisfies the music classification condition, classify the current audio frame as a music frame.

16. The apparatus according to any one of claims 9 to 14, further comprising:

a parameter obtaining unit, configured to obtain a frequency spectrum high-frequency-band peakiness, a frequency spectrum correlation degree, a voicing parameter, and a linear prediction residual energy tilt of the current audio frame, wherein the frequency spectrum high-frequency-band peakiness denotes a peakiness or an energy acutance, on a high frequency band, of a frequency spectrum of the current audio frame; the frequency spectrum correlation degree denotes stability, between adjacent frames, of a signal harmonic structure of the current audio frame; the voicing parameter denotes a time domain correlation degree between the current audio frame and a signal before a pitch period; and the linear prediction residual energy tilt denotes an extent to which linear prediction residual energy of the audio signal changes as a linear prediction order increases; wherein

the storage determining unit is further configured to determine, according to the voice activity of the current audio frame, whether to store the frequency spectrum high-frequency-band peakiness, the frequency spectrum correlation degree, and the linear prediction residual energy tilt in memories;

the memory is further configured to: when the storage determining unit outputs a result that the frequency spectrum high-frequency-band peakiness, the frequency spectrum correlation degree, and the linear prediction residual energy tilt need to be stored, store the frequency spectrum high-frequency-band peakiness, the frequency spectrum correlation degree, and the linear prediction residual energy tilt; and

the classification unit is specifically configured to obtain statistics of effective data of the stored frequency spectrum fluctuations, statistics of effective data of stored frequency spectrum high-frequency-band peakiness, statistics of effective data of stored frequency spectrum correlation degrees, and statistics of effective data of stored linear prediction residual energy tilts, and classify the audio frame as a speech frame or a music frame according to the statistics of the effective data.

17. The apparatus according to claim 16, wherein the classification unit comprises:

a calculating unit, configured to obtain an average value of the effective data of the stored frequency spectrum fluctuations, an average value of the effective data of the stored frequency spectrum high-frequency-band peakiness, an average value of the effective data of the stored frequency spectrum correlation degrees, and a variance of the effective data of the stored linear prediction residual energy tilts separately; and

a determining unit, configured to: when a classifying condition comprising one of the following conditions is satisfied, classify the current audio frame as a music frame; otherwise classify the current audio frame as a speech frame: the average value of the effective data of the frequency spectrum fluctuations is less than a first threshold; or the average value of the effective data of the frequency spectrum high-frequency-band peakiness is greater than a second threshold; or the average value of the effective data of the frequency spectrum correlation degrees is greater than a third threshold; or the variance of the effective data of the linear prediction residual energy tilts is less than a fourth threshold.