	Previous N^{th} frame	...	Previous second frame	Previous frame	Current frame	

FIG. 1

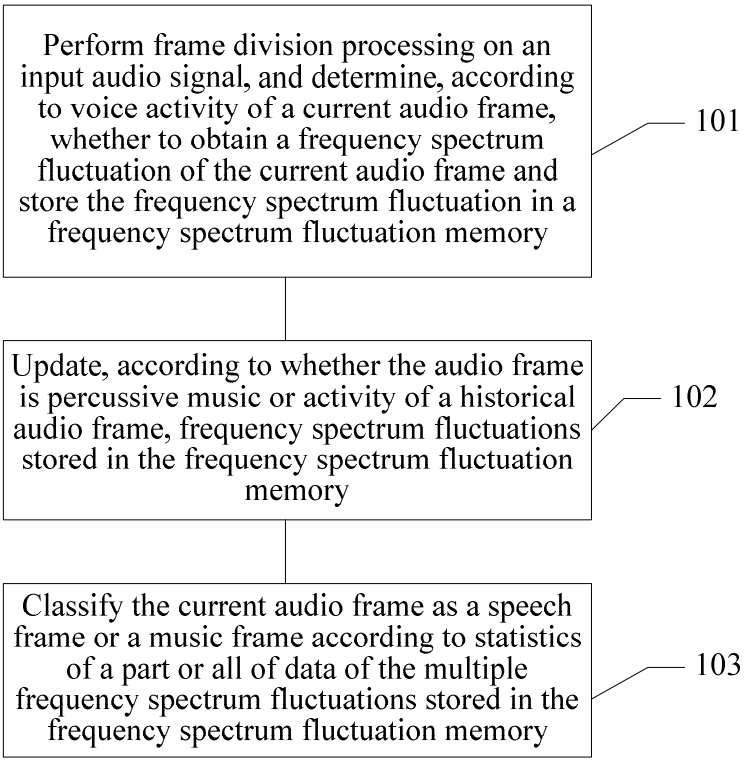


FIG. 2

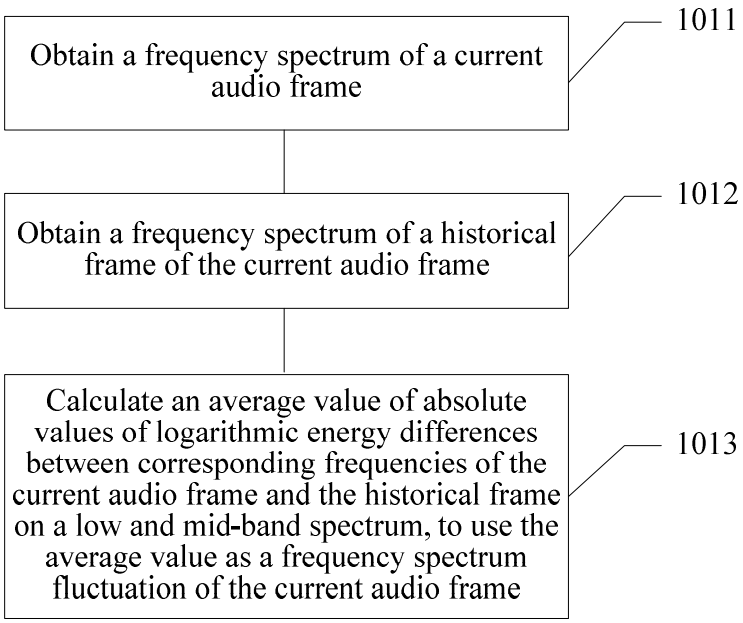

Obtain a frequency spectrum of a current audio frame
1011
Obtain a frequency spectrum of a historical frame of the current audio frame
1012
Calculate an average value of absolute values of logarithmic energy differences between corresponding frequencies of the current audio frame and the historical frame on a low and mid-band spectrum, to use the average value as a frequency spectrum fluctuation of the current audio frame
1013

FIG. 3

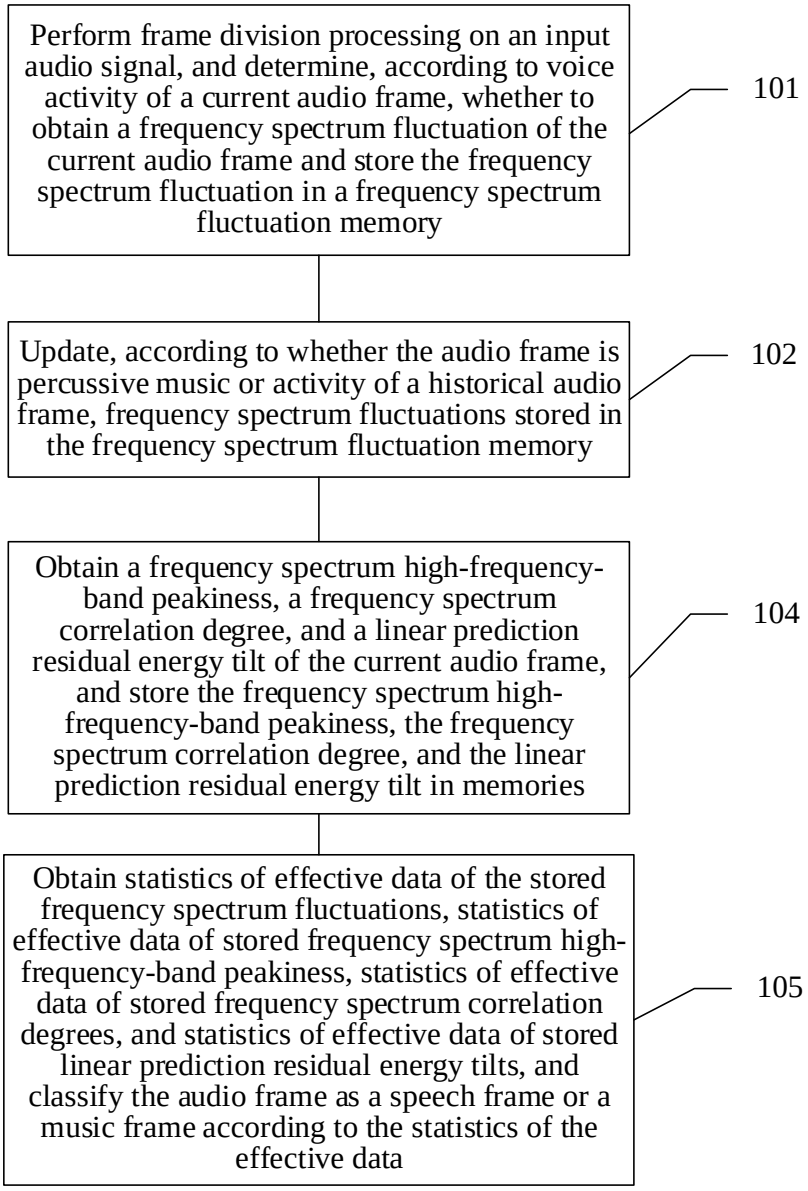

Perform frame division processing on an input audio signal, and determine, according to voice activity of a current audio frame, whether to obtain a frequency spectrum fluctuation of the current audio frame and store the frequency spectrum fluctuation in a frequency spectrum fluctuation memory
101
Update, 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
102
Obtain a frequency spectrum high-frequency-band peakiness, a frequency spectrum correlation degree, and a linear prediction residual energy tilt of the current audio frame, and store the frequency spectrum high-frequency-band peakiness, the frequency spectrum correlation degree, and the linear prediction residual energy tilt in memories
104
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
105

FIG. 4

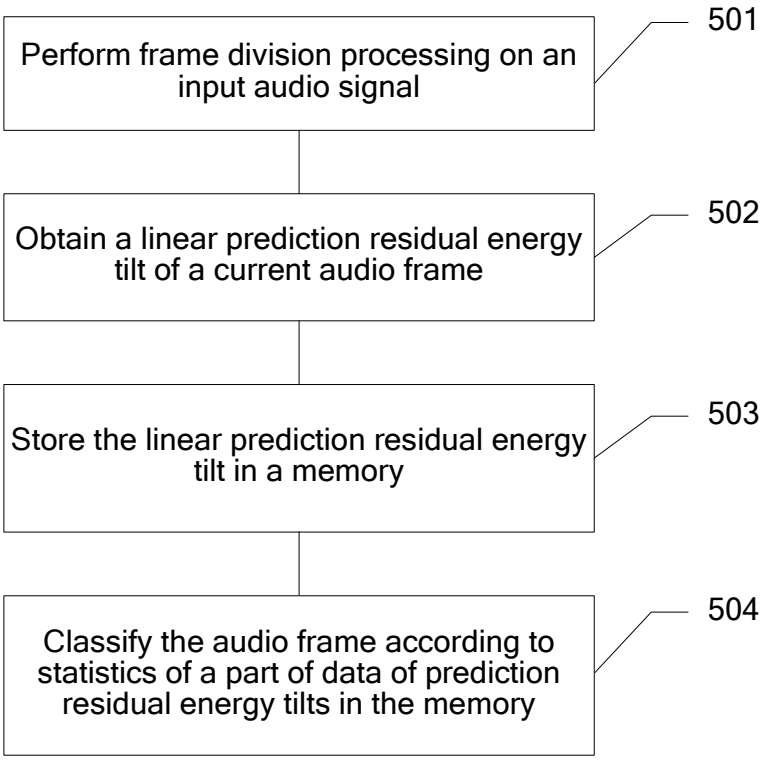

Perform frame division processing on an input audio signal
501
Obtain a linear prediction residual energy tilt of a current audio frame
502
Store the linear prediction residual energy tilt in a memory
503
Classify the audio frame according to statistics of a part of data of prediction residual energy tilts in the memory
504

FIG. 5

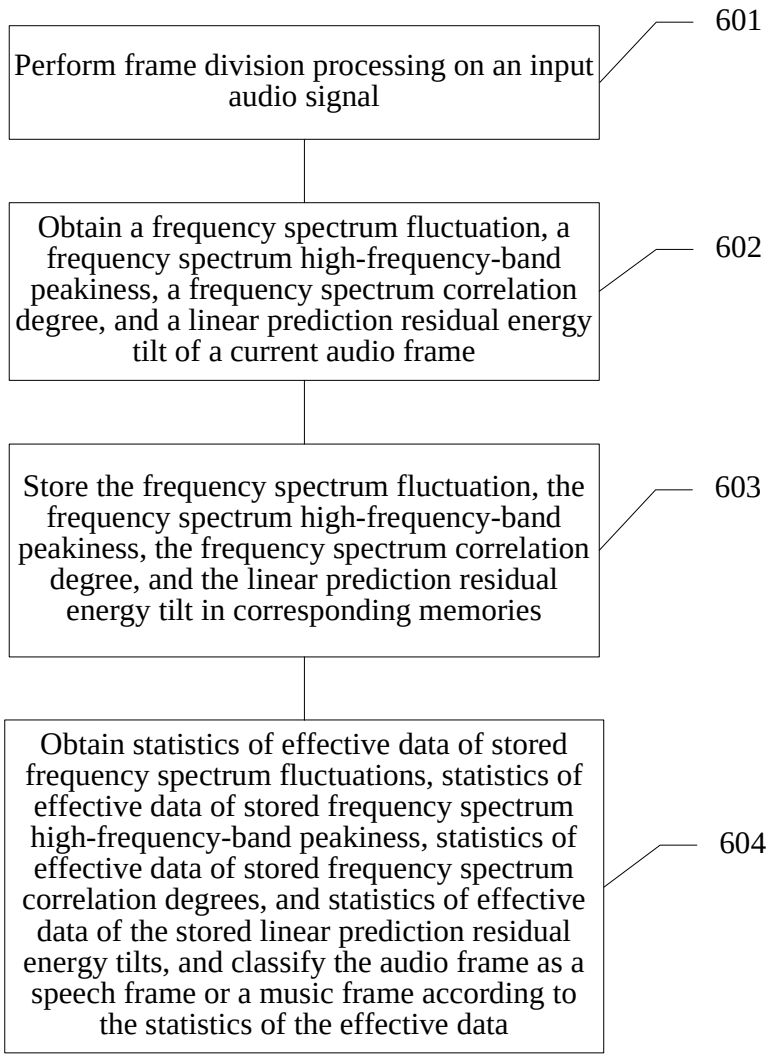

Perform frame division processing on an input audio signal
601
Obtain a frequency spectrum fluctuation, a frequency spectrum high-frequency-band peakiness, a frequency spectrum correlation degree, and a linear prediction residual energy tilt of a current audio frame
602
Store the frequency spectrum fluctuation, the frequency spectrum high-frequency-band peakiness, the frequency spectrum correlation degree, and the linear prediction residual energy tilt in corresponding memories
603
Obtain statistics of effective data of 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 the 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
604

FIG. 6

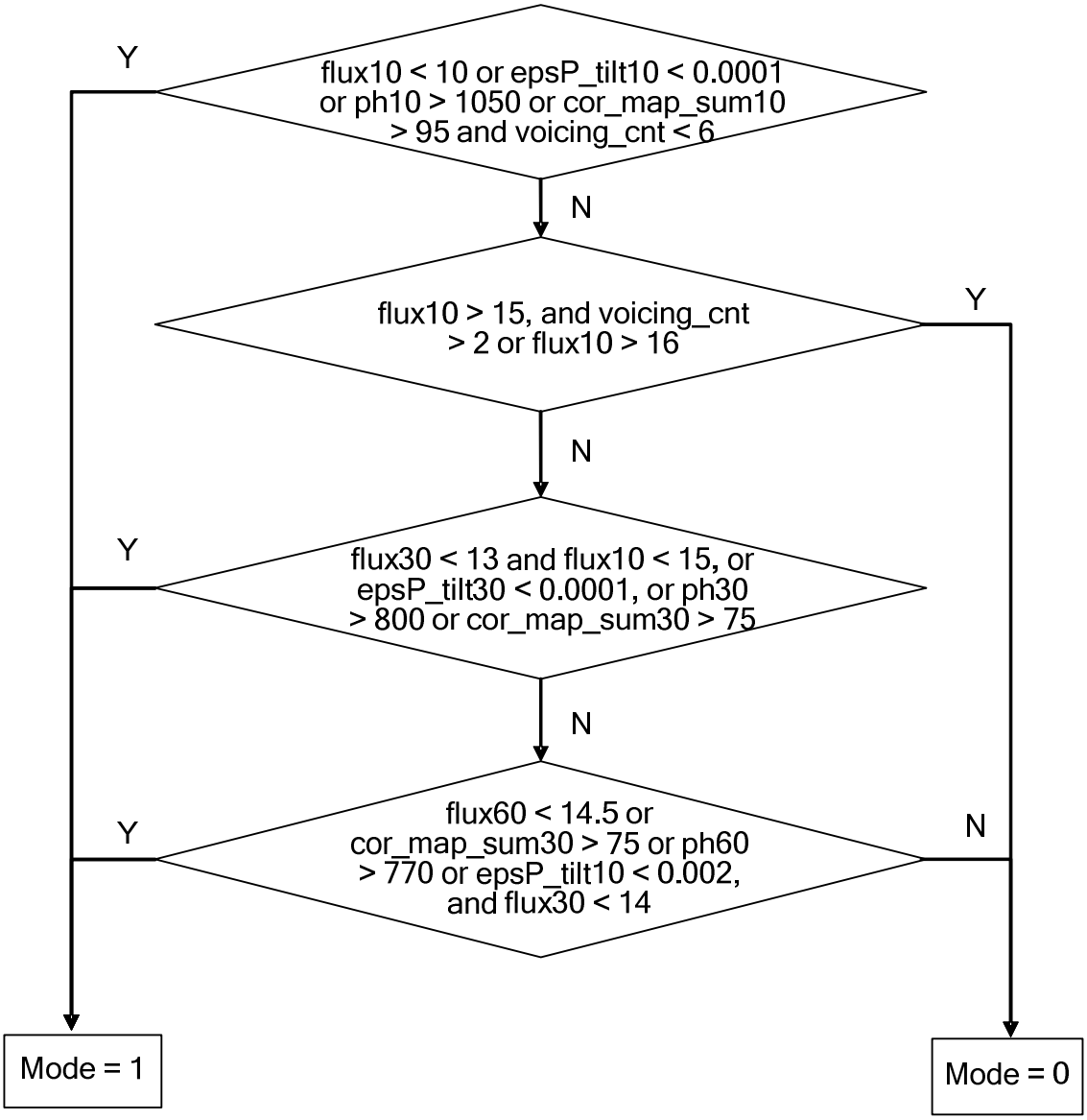

flux10 < 10 or epsP_tilt10 < 0.0001 or ph10 > 1050 or cor_map_sum10 > 95 and voicing_cnt < 6
Y
N
flux10 > 15, and voicing_cnt > 2 or flux10 > 16
Y
N
flux30 < 13 and flux10 < 15, or epsP_tilt30 < 0.0001, or ph30 > 800 or cor_map_sum30 > 75
Y
N
flux60 < 14.5 or cor_map_sum30 > 75 or ph60 > 770 or epsP_tilt10 < 0.002, and flux30 < 14
Y
N
Mode = 1
Mode = 0

FIG. 7

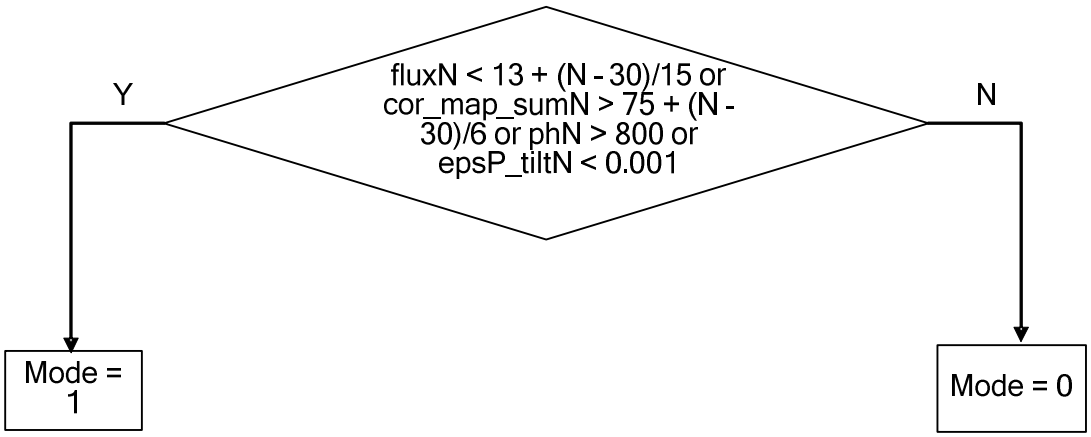

fluxN < 13 + (N - 30)/15 or cor_map_sumN > 75 + (N - 30)/6 or phN > 800 or epsP_tiltN < 0.001
Y
N
Mode = 1
Mode = 0

FIG. 8

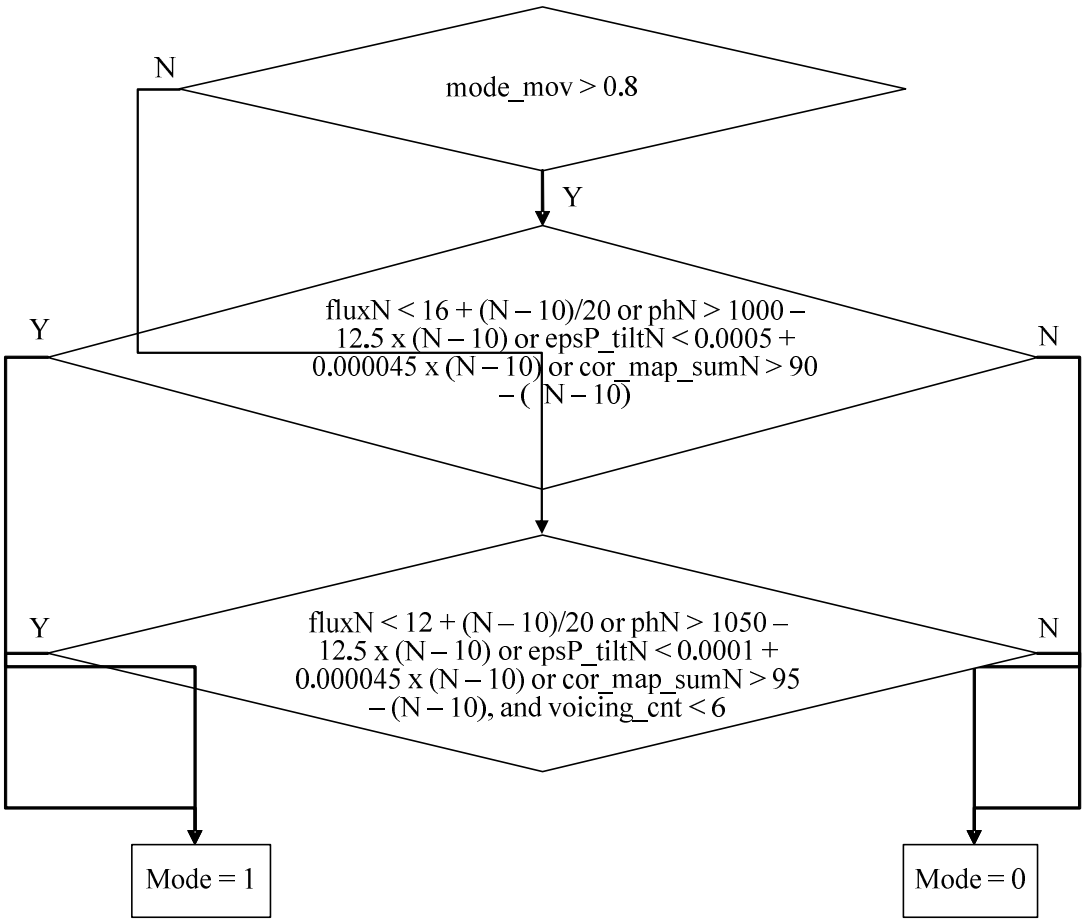

mode_mov > 0.8
N
Y
fluxN < 16 + (N – 10)/20 or phN > 1000 – 12.5 x (N – 10) or epsP_tiltN < 0.0005 + 0.000045 x (N – 10) or cor_map_sumN > 90 – (N – 10)
Y
N
fluxN < 12 + (N – 10)/20 or phN > 1050 – 12.5 x (N – 10) or epsP_tiltN < 0.0001 + 0.000045 x (N – 10) or cor_map_sumN > 95 – (N – 10), and voicing_cnt < 6
Y
N
Mode = 1
Mode = 0

FIG. 9

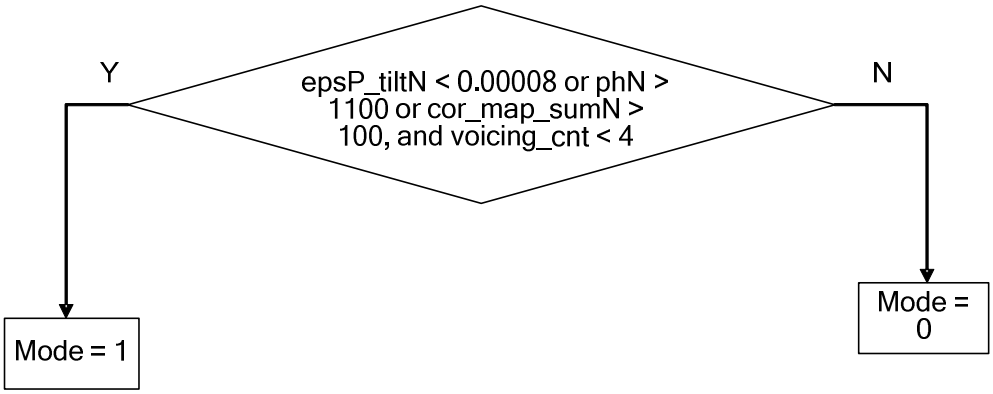

epsP_tiltN < 0.00008 or phN > 1100 or cor_map_sumN > 100, and voicing_cnt < 4
Y
N
Mode = 1
Mode = 0

FIG. 10

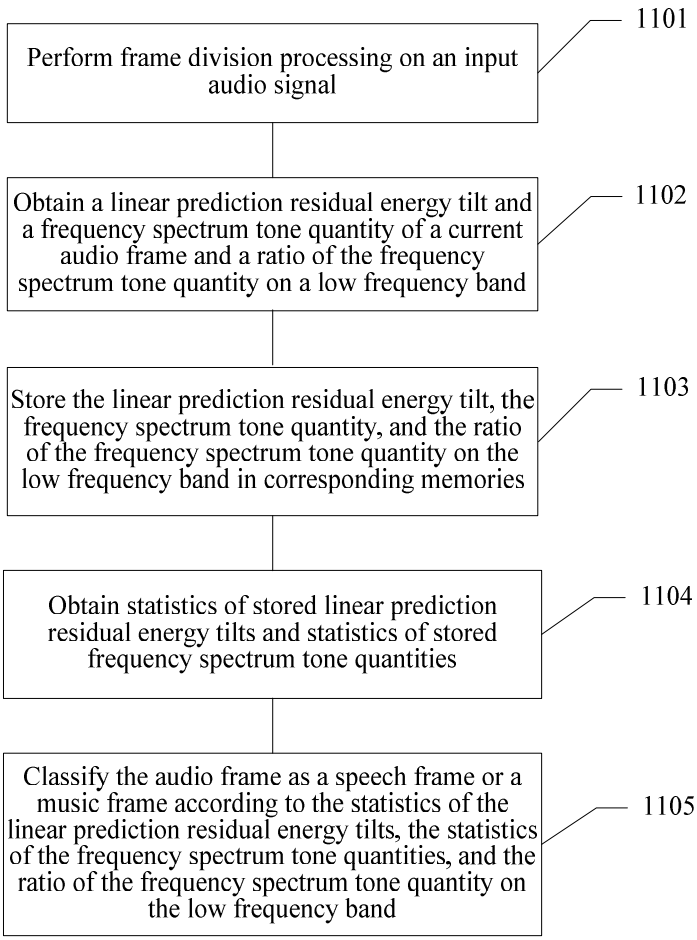

Perform frame division processing on an input audio signal
1101
Obtain a linear prediction residual energy tilt and a frequency spectrum tone quantity of a current audio frame and a ratio of the frequency spectrum tone quantity on a low frequency band
1102
Store the linear prediction residual energy tilt, the frequency spectrum tone quantity, and the ratio of the frequency spectrum tone quantity on the low frequency band in corresponding memories
1103
Obtain statistics of stored linear prediction residual energy tilts and statistics of stored frequency spectrum tone quantities
1104
Classify the audio frame as a speech frame or a music frame according to the statistics of the linear prediction residual energy tilts, the statistics of the frequency spectrum tone quantities, and the ratio of the frequency spectrum tone quantity on the low frequency band
1105

FIG. 11

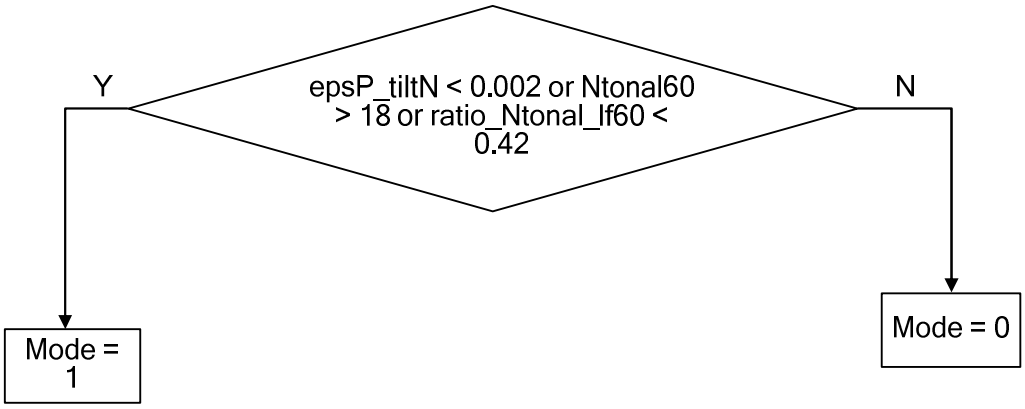


FIG. 12

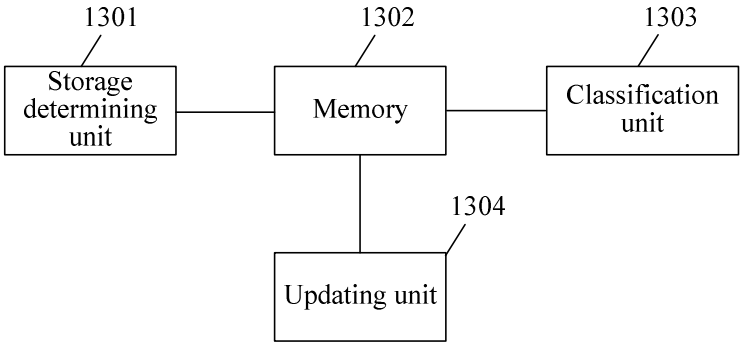


FIG. 13

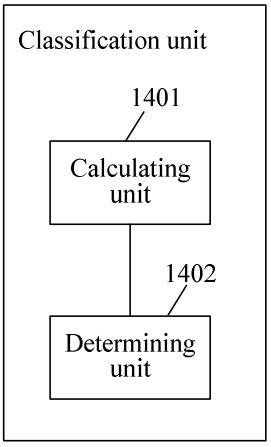


FIG. 14

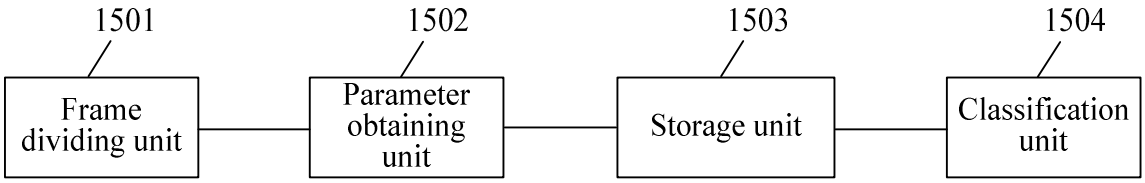


FIG. 15

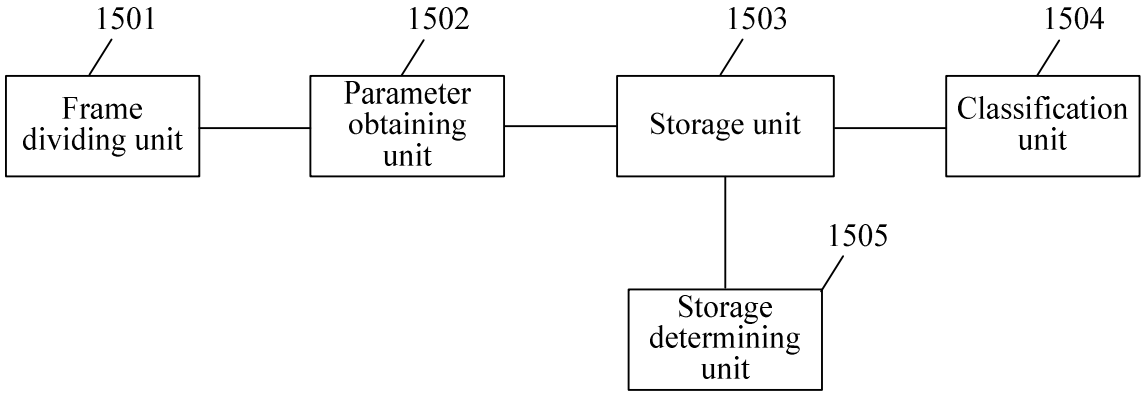


FIG. 16

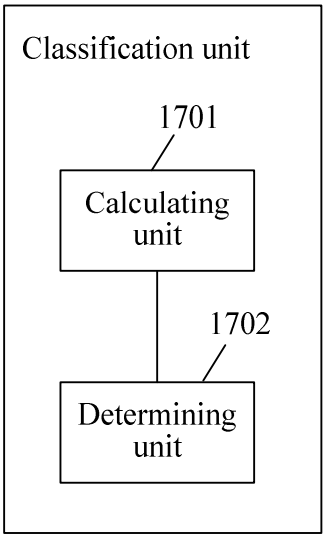


FIG. 17

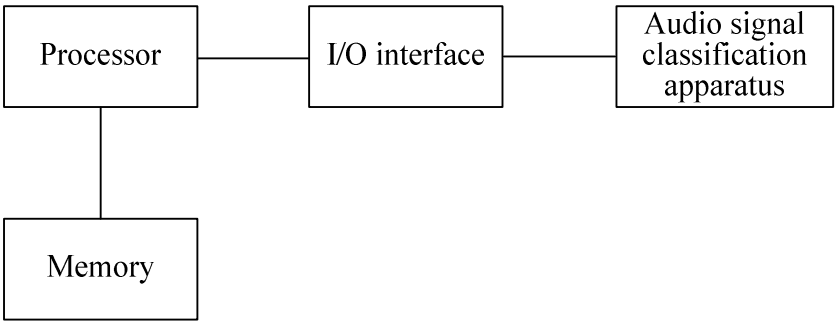

Processor
I/O interface
Audio signal classification apparatus
Memory

FIG. 18

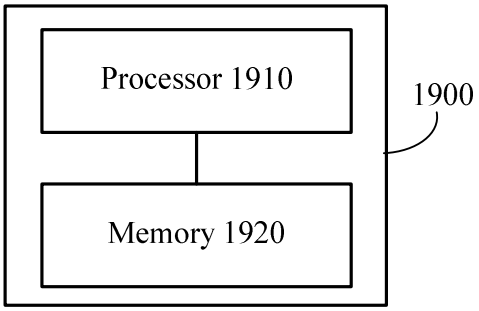

Processor 1910
1900
Memory 1920

FIG. 19