

Method and Apparatus for Controlling Video-audio Data Playing

Field of the Invention

The present invention relates to network communication technologies, and more particularly to a method and apparatus for controlling video-audio data playing.

Background of the Invention

With the development of Internet, more and more users communicate with each other on the Internet. With users' further requirements for communication and the further development of the Internet, the users can communicate with each other on the Internet not only through characters but also through video-audio data.

In the prior art, since data transmission on the Internet is based on a Transmission Control Protocol/Internet Protocol (TCP/IP), when video-audio data are transmitted on the Internet, it is needed to divide the video-audio data into multiple data packets according to the TCP/IP and transmit the multiple data packets on the Internet by taking a frame as a unit. Since the network structure of the Internet is complex, the transmission rate of the data packets transmitted on the Internet can not be controlled effectively, and a receiving end of the video-audio data sometimes can not receive the video-audio data for a long time, which results in disconnection or incontinuity phenomena when the video-audio data is recovered and played, e.g. audio incontinuity, call mute, video image standstill and so on.

In order to solve the problem, a buffer is usually configured on the receiving end of the video-audio data to store some video-audio data, and the video-audio data are received from the network, and then decoded and to be played, so as to decrease the above phenomena. But the capacity of the buffer is limited, if the buffer does not receive subsequent video-audio data after the video-audio data in the buffer has been played, the incontinuity phenomena will occur when video-audio data is played, which decreases user experiences.

Summary of the Invention

Embodiments of the present invention provide a method and apparatus for controlling video-audio data playing, so as to decrease the incontinuity phenomena when video-audio data is played and improve user experiences.

The technical solutions provided by the embodiments of the present invention are
5 implemented as follows.

A method for controlling video-audio data playing includes:

obtaining total duration of playing video-audio data which can be played in a buffer;

calculating a time difference, the time difference being a difference between a
current time point and a time point of receiving a latest video-audio data; and

10 comparing the time difference with a preset condition, performing time domain
stretch processing for the video-audio data in the buffer according to a comparing result,
and playing the processed video-audio data.

An apparatus for controlling video-audio data playing includes:

an obtaining unit, configured to obtain total duration of playing video-audio data
15 which can be played in a buffer;

a calculating unit, configured to calculate a time difference, the time difference being
a difference between a current time point and a time point of receiving a latest
video-audio data packet; and

a processing unit, configured to compare the time difference with a preset condition,
20 perform time domain stretch processing for the video-audio data in the buffer according
to a comparing result, and play the processed video-audio data.

As can be seen from the above technical solutions, in the embodiments of the present
invention, by comparing the time difference with a preset condition and performing time
domain stretch processing for the video-audio data according to a comparing result, the
25 total duration of playing video-audio data in the buffer is increased, and when the
video-audio data is played, the incontinuity phenomena caused because the receiving end
can not receive a data packet for a long time is decreased, so as to improve user
experiences.

Brief Description of the Drawings

Figure 1 is a flowchart illustrating a method for controlling video-audio data playing according to an embodiment of the present invention.

Figure 2 is a schematic diagram illustrating the structure of an apparatus for controlling video-audio data playing according to an embodiment of the present invention.

Figure 3 is a schematic diagram illustrating the processing of a playing control procedure according to a first embodiment of the present invention.

Figure 4 is a schematic diagram illustrating the processing of a playing control procedure according to a second embodiment of the present invention.

10 **Detailed Description of the Invention**

In order to make the object, technical solution and merits of the present invention clearer, the present invention will be illustrated in detail hereinafter with reference to the accompanying drawings and embodiments.

In the embodiments of the present invention, according to a short-term correlation between audio signals and by using a Waveform Similarity Overlap and Add (WSOLA) algorithm, one segment of audio signals is divided into shorter wave bands, and the wave bands are superimposed and combined to obtain a signal with a required playing duration. If the duration of playing the processed audio signals is longer than the duration of playing the original audio signals, which is equivalent to performing time domain stretch processing for the original audio signals, the duration of playing the original audio signals is prolonged. If the duration of playing the processed audio signals is shorter than the duration of playing the original audio signals, which is equivalent to performing time domain compacting processing for the original audio signals, the duration of playing the original audio signals is shortened.

Based on this idea, the embodiments of the present invention provide a playing control method to self-adaptively control the playing rate, so as to adapt to the discontinuity of receiving packets in the network. Specifically, when it is detected that no data are received for a long time, time domain stretch processing is performed for data to be played, so as to increase the duration of playing the data, thereby solving the playing

incontinuity problem; further, when it is detected that the receiving rate of network data is too high, time domain compacting processing is performed for the data to be played, so as to decrease the duration of playing the data, thereby solving both the playing incontinuity problem and the delay problem. The time domain processing method only changes the playing rate, but does not influence the semanteme, and thus does not further influence user experiences.

The embodiments of the present invention provide a method for controlling video-audio data playing, as shown in Figure 1. The method includes following steps.

Step 11: total duration of playing video-audio data which can be played in a buffer is obtained.

In this step, the mode of obtaining the total duration of playing the video-audio data includes: obtaining the total duration of playing the video-audio data which can be played in the buffer in one process, or periodically obtaining the total duration of playing the video-audio data from the buffer. The embodiments of the present invention do not limit the obtaining times and mode.

Step 12: a time difference is calculated, and the time difference is a difference between a current time point and a time point of receiving the latest video-audio data packet.

Step 13: the time difference is compared with a preset condition, time domain stretch processing is performed for the video-audio data in the buffer according to a comparing result, and the processed video-audio data is played.

The time domain stretch processing method in Step 13 includes:

when determining that the time difference is larger than or equal to the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to preset minimum duration for the buffer, performing time domain stretch processing for the video-audio data in the buffer; otherwise, not performing the time domain stretch processing. The duration of playing the video-audio data for which the time domain stretch processing is performed is not larger than two times of the duration of playing the original

video-audio data. In actual applications, the times can be adjusted, e.g. 1.5 times, 2.5 times or 3 times, which is not limited by the embodiments of the present invention.

In addition, when the time domain stretch processing is performed for the video-audio data in the buffer, a determining condition is configured, i.e. when it is
5 determined that the time difference is larger than or equal to the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to the preset minimum duration for the buffer, further determining whether a cumulative delay of the video-audio data is smaller than a preset maximum cumulative delay of video-audio data;
10 when the cumulative delay of the video-audio data is smaller than the preset maximum cumulative delay of the video-audio data, performing time domain stretch processing for the video-audio data in the buffer; otherwise, not performing time domain stretch processing. In the method, the condition of performing time domain stretch processing is limited to a case that the cumulative delay of the video-audio data is smaller than the
15 preset maximum cumulative delay of the video-audio data, so as to avoid that overmuch time domain stretch processing results in the infinite increase of the cumulative delay of the video-audio data, and further improve user experiences.

An stretch or compacting algorithm used in the stretch or compacting processing includes: for an audio data file, an algorithm such as a Synchronized Overlap Add (SOLA)
20 algorithm and a Waveform Similarity Overlap and Add (WSOLA) algorithm, and some improved algorithms based on this type of algorithms are used to implement the stretch or compacting processing; for a video data file, a repeat inserting frame method is used, or a method in which adjacent two frames are combined into a middle frame is used to implement the stretch processing, and a video skip frame method is used to implement the
25 compacting processing. The embodiments of the present invention do not limit the algorithm for implementing the stretch or compacting of the video-audio data file. The times of stretch or compacting of the video-audio data file may be configured by users, and is not limited by the embodiments of the present invention.

In addition, based on the above determining of the stretch processing, if the
30 determining result is not to perform stretch processing, further determining is performed. If it is determined that the time difference is smaller than the total duration of playing the

video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is larger than preset maximum duration for the buffer, the time domain compacting processing is performed for the video-audio data in the buffer; otherwise, the time domain compacting processing is not performed. The duration of playing the vide-audio data for which the time domain compacting processing is performed is not smaller than $1/2$ times of the duration of playing the original video-audio data. In actual applications, the times can be adjusted, e.g. $1/3$ times, $2/3$ times or $3/4$ times, which is not limited by the embodiments of the present invention.

10 By performing the time domain compacting processing for the video-audio data in the buffer, the delay of playing the video-audio data can be decreased, so as to further improve user experiences, and make a receiving end adaptively play the video-audio data according to the receiving state of the video-audio data.

The method for controlling video-audio data playing is described hereinafter in detail by taking audio data as an example.

First, the following variables are configured:

Jpre: a time difference between a time point of receiving the latest audio data packet and a current time point;

Lp: duration of playing audio data which can be played in the buffer;

20 Dalg: a cumulative algorithm delay, i.e. a cumulative value of playing delay caused by the time domain stretch and compacting processing;

Dmax: the upper limit of the cumulative algorithm delay, i.e. a maximum value that Dalg can achieve;

Lmin: the lower limit of the duration of playing audio data for the buffer;

25 Lmax: the upper limit of the duration of playing audio data for the buffer;

Lout: duration of playing audio data which is taken away from the buffer each time;

Lstretch: duration increased by the stretch processing;

Lcompact: duration decreased by the compacting processing;

The unit of the above duration of playing audio data or time is millisecond (ms).

Figure 3 is a schematic diagram illustrating the processing of a playing control
5 procedure according to a first embodiment of the present invention. As shown in Figure 3,
the playing control procedure includes the following steps.

Step 301: audio data with a certain quantity is obtained and played periodically, and
Lp and Jpre are updated, where the updated $L_p = L_p \text{ before the update} - L_{out}$, and $J_{pre} =$
a current time point - a time point of receiving the latest audio packet, the time point of
10 receiving the latest audio packet is updated every time one audio packet is received.

Step 302: it is determined whether $J_{pre} \geq L_p$ and $L_p \leq L_{min}$; if $J_{pre} \geq L_p$ or L_p
 $\leq L_{min}$, Step 304 is performed; otherwise, Step 303 is performed.

Step 303: it is determined whether $J_{pre} < L_p$ and $L_p > L_{max}$; if $J_{pre} < L_p$ or $L_p >$
 L_{max} , Step 305 is performed; otherwise, Step 301 is performed.

15 Step 304: time domain stretch processing is performed for the audio data in the
buffer, the processed audio data is stored in the buffer, Lp and Dalg are updated, where
the updated $L_p = L_p \text{ before the update} + L_{stretch}$, and the updated $D_{alg} = D_{alg} \text{ before the}$
update + Lstretch, and Step 301 is returned to.

Step 305: time domain compacting processing is performed for the audio data in the
20 buffer, the processed audio data is stored in the buffer, Lp and Dalg are updated, where
the updated $L_p = L_p \text{ before the update} - L_{compact}$, and the updated $D_{alg} = D_{alg} \text{ before}$
the update- Lcompact, and Step 301 is returned to.

In the above method, Lmin may be configured as different values according to
specific applications, generally $L_{min} \geq 3 * L_{out}$. Since there may be no data in the
25 buffer at the very start, the data in the buffer can not be played unless the duration of
playing the data in the buffer is equal to Lmin, and thus some playing delay is increased.

Lmax may be determined according to the size of the buffer in specific applications,

as long as a basic condition is met: $L_{\max} > L_{\min}$.

L_{stretch} : duration increased by the stretch processing, where a better effect can be achieved when the duration is not larger than the duration of playing the original data, and of course the duration may be determined according to practical requirements.

5 L_{compact} : duration decreased by the compacting processing, where a better effect can be achieved when the duration is not larger than 1/2 times of the duration of playing the original data, and of course the duration may be determined according to practical requirements.

10 In the above playing control method, $L_{\min}$ will introduce a constant playing delay, which increases a fixed playing delay. The time domain stretch processing also introduces a delay, which is determined by the duration increased by each stretch processing, and increases the playing delay. The time domain compacting processing decreases a delay, which is determined by the duration decreased by each compacting processing, and decreases the playing delay. The total playing delay may be decreased by the time domain
15 compacting processing, so as to counteract the playing delay introduced by $L_{\min}$ and the stretch processing.

20 In order to control the delay, the cumulative delay may be taken as a condition in the control algorithm. When the stretch processing is performed, D_{alg} is considered; if $D_{\text{alg}} \geq D_{\max}$, the stretch processing is not performed, to avoid that a larger playing delay is introduced.

Figure 4 is a schematic diagram illustrating the processing of a playing control procedure according to a second embodiment of the present invention. As shown in Figure 4, the playing control procedure includes the following steps.

25 Step 401: data with a certain quantity is obtained and played, and L_p and J_{pre} are updated.

Step 402: it is determined whether $J_{\text{pre}} \geq L_p$ and $L_p \leq L_{\min}$; if $J_{\text{pre}} \geq L_p$ or $L_p \leq L_{\min}$, Step 404 is performed; otherwise, Step 403 is performed.

Step 403: it is determined whether $J_{\text{pre}} < L_p$, $L_p > L_{\max}$ and $D_{\text{alg}} > D_{\max}$; if $J_{\text{pre}} <$

Lp, or $L_p > L_{max}$, or $D_{alg} > D_{max}$, Step 405 is performed; otherwise, Step 401 is performed.

Step 404: it is determined whether $D_{alg} > D_{max}$; if $D_{alg} > D_{max}$, Step 407 is performed; otherwise, Step 406 is performed.

5 Step 405: time domain compacting processing is performed for the data in the buffer, the processed data is stored in the buffer, Lp and Dalg are updated, and Step 401 is returned to.

10 Step 406: time domain stretch processing is performed for the data in the buffer, the processed data is stored in the buffer, Lp and Dalg are updated, and Step 401 is returned to.

Step 407: according to the data of the former frame, data with the certain quantity is obtained, the data is stored in the buffer, Lp and Dalg are updated, and Step 401 is returned to.

15 In the above two specific embodiments, the step of determining whether the compacting processing needs to be performed is not necessary, if the total playing delay does not need to be controlled, the step of determining whether the compacting processing needs to be performed and the step of performing the compacting processing are not necessary, and Step 403 may not include the step of determining whether $D_{alg} > D_{max}$.

20 The embodiments of the present invention also provide an apparatus for controlling video-audio data playing, as shown in Figure 2. The apparatus includes: an obtaining unit 21, configured to obtain total duration of playing video-audio data which can be played in the buffer; a calculating unit 22, configured to calculate a time difference which is a difference between a current time point and a time point of receiving the latest
25 video-audio data packet; a processing unit 23, configured to compare the time difference with a preset condition, perform time domain stretch processing for the video-audio data in the buffer, and play the processed video-audio data.

Preferably, the processing unit 23 may include: a buffer determining module 231, configured to determine whether the time difference is larger than or equal to the total

duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to preset minimum duration for the buffer; an stretch module 232, configured to, when the determining result of the buffer determining module 231 is yes, i.e. when the
5 buffer determining module 231 determines that the time difference is larger than or equal to the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to the preset minimum duration for the buffer, perform the time domain stretch processing for the video-audio data in the buffer, and play the video-audio
10 data after the time domain stretch processing.

Preferably, the stretch module 232 may further include: an stretch sub-module 2321, configured to determine whether a cumulative delay of the video-audio data is smaller than the preset maximum cumulative delay of the video-audio data; the stretch module 232 is further configured to, when the determining result of the buffer determining
15 module 231 is yes, determine whether to perform time domain stretch processing for the video-audio data in the buffer according to the determining result of the stretch sub-module 2321, when the determining result of the stretch sub-module 2321 is yes, i.e. when the stretch sub-module 2321 determines that the cumulative delay of the video-audio data is smaller than the preset maximum cumulative delay of the video-audio
20 data, perform the time domain stretch processing for the video-audio data in the buffer; otherwise, not perform the time domain stretch processing; the stretch module 232 may further include a playing sub-module 2322, configured to play the video-audio data after the time domain stretch processing.

Preferably, the processing unit 23 may further include: a compacting determining
25 module 233 and a compacting module 234. The compacting determining module 233 is configured to determine whether the time difference is smaller than the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is larger than preset maximum duration for the buffer.

30 The compacting module 234 is configured to, when the determining result of the compacting determining module 233 is yes, i.e. when the compacting determining module

233 determines that the time difference is smaller than the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is larger than the preset maximum duration for the buffer, perform time domain compacting processing for the video-audio data in the buffer, and play the video-audio data after the time domain compacting processing.

Preferably, the compacting determining module 233 is further configured to determine whether the cumulative delay of the video-audio data is smaller than or equal to the preset maximum cumulative delay of the video-audio data;

the compress module 234, further configured to, when the compacting determining module 233 determines that the cumulative delay of the video-audio data is smaller than or equal to the preset maximum cumulative delay of the video-audio data, perform time domain compacting processing for the video-audio data in the buffer.

According to the apparatus provided by the embodiments of the present invention, after the obtaining unit 21 obtains the total duration of playing the video-audio data which can be played in the buffer and the calculating unit 22 calculates the time difference, the buffer determining module 231 determines whether the time difference is larger than or equal to the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to the preset minimum duration for the buffer; if the determining result of the buffer determining module 231 is yes, the stretch module 232 performs the time domain stretch processing for the video-audio data in the buffer, and plays the video-audio data after the time domain stretch processing. By using the apparatus, the duration of playing the video-audio data in the buffer is increased, which increases the capacity of the buffer without increasing the size of the buffer, so as to provide more time for receiving other video-audio data packets by the receiving end, and decrease the discontinuity phenomena caused because the receiving end can not receive a data packet for a long time when the video-audio data is played. By using the apparatus, when the buffer determining module 231 determines that the time difference is smaller than the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is

larger than the preset maximum duration for the buffer, the compacting module 234 performs the time domain compacting processing for the video-audio data in the buffer, and plays the video-audio data after the time domain compacting processing, so as to decrease the playing delay of video-audio data, improve user experiences, and adaptively
5 play the video-audio data.

Those skilled in the art can understand that all or part of steps in the method provided by the embodiments of the present invention can be implemented by instructing related hardware by a program, the program may be stored a readable memory of a computer, and above method steps are included when the program is operated. The
10 memory includes a ROM/RAM, a disk, a Compact Disc (CD) and so on.

To sum up, the technical solution provided by the embodiments of the present invention has an advantage of adaptively playing a data file.

The foregoing are only preferred embodiments of the present invention and are not for use in limiting the protection scope of the present invention. Any modification,
15 equivalent replacement and improvement made within the scope of the present invention should be covered under the protection scope of the present invention. Therefore, the protection scope of the present invention should be defines according to the claims.

Claims

What is claimed is:

1. A method for controlling video-audio data playing, comprising:

obtaining total duration of playing video-audio data which can be played in a buffer;

5 calculating a time difference, the time difference being a difference between a current time point and a time point of receiving a latest video-audio data; and

comparing the time difference with a preset condition, performing time domain stretch processing for the video-audio data in the buffer according to a comparing result, and playing the processed video-audio data.

10

2. The method of claim 1, wherein comparing the time difference with a preset condition and performing time domain stretch processing for the video-audio data in the buffer according to a comparing result comprises:

when determining that the time difference is larger than or equal to total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to preset minimum duration for the buffer, performing the time domain stretch processing for the video-audio data in the buffer.

20 3. The method of claim 2, after determining that the time difference is larger than or equal to the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to the preset minimum duration for the buffer, and before performing the time domain stretch processing for the video-audio data in the buffer, further comprising:

25

determining whether a cumulative delay of the video-audio data is smaller than a preset maximum cumulative delay of video-audio data;

when the cumulative delay of the video-audio data is smaller than the preset maximum cumulative delay of the video-audio data, performing the time domain stretch processing for the video-audio data in the buffer.

30

4. The method of claim 2 or 3, if it is determined that it is not needed to perform the

time domain stretch processing, further comprising:

when determining that the time difference is smaller than the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is larger than preset maximum duration for the buffer, performing time domain compacting processing for the video-audio data in the buffer.

5. The method of claim 4, if it is determined that it is not needed to perform time domain stretch processing, further comprising:

when determining that the cumulative delay of the video-audio data is smaller than or equal to the preset maximum cumulative delay of the video-audio data, performing time domain compacting processing for the video-audio data in the buffer.

6. An apparatus for controlling video-audio data playing, comprising:

an obtaining unit, configured to obtain total duration of playing video-audio data which can be played in a buffer;

a calculating unit, configured to calculate a time difference, the time difference being a difference between a current time point and a time point of receiving a latest video-audio data packet; and

a processing unit, configured to compare the time difference with a preset condition, perform time domain stretch processing for the video-audio data in the buffer according to a comparing result, and play the processed video-audio data.

7. The apparatus of claim 6, wherein the processing unit comprises:

a buffer determining module, configured to determine whether the time difference is larger than or equal to the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to preset minimum duration for the buffer;

an stretch module, configured to, when the buffer determining module determines that the time difference is larger than or equal to the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to the preset minimum duration for the buffer, perform the time domain stretch processing for the

video-audio data in the buffer, and play the video-audio data after the time domain stretch processing.

8. The apparatus of claim 7, wherein the stretch module further comprises:

5 an stretch sub-module, configured to determine whether a cumulative delay of the video-audio data is smaller than a preset maximum cumulative delay of the video-audio data; wherein

the stretch module is further configured to, when the determining module determines that the time difference is larger than or equal to the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is smaller than or equal to the preset minimum duration for the buffer, and the stretch sub-module determines that the cumulative delay of the video-audio data is smaller than the preset maximum cumulative delay of the video-audio data, perform the time domain stretch processing for the video-audio data in the buffer; and

15 a playing sub-module, configured to play the video-audio data after the time domain stretch processing.

9. The apparatus of claim 7 or 8, wherein the processing unit further comprises:

20 a compacting determining module, configured to determine whether the time difference is smaller than the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is larger than preset maximum duration for the buffer; and

a compacting module, configured to, when the compacting determining module determines that the time difference is smaller than the total duration of playing the video-audio data which can be played in the buffer, or the total duration of playing the video-audio data which can be played in the buffer is larger than the preset maximum duration for the buffer, perform time domain compacting processing for the video-audio data in the buffer, and play the video-audio data after the time domain compacting processing.

10. The apparatus of claim 9, wherein

the compacting determining module, is further configured to determine whether the

cumulative delay of the video-audio data is smaller than or equal to the preset maximum cumulative delay of the video-audio data;

the compacting module, further configured to, when the compacting determining module determines that the cumulative delay of the video-audio data is smaller than or
5 equal to the preset maximum cumulative delay of the video-audio data, perform time domain compacting processing for the video-audio data in the buffer.