

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3629253号
(P3629253)

(45) 発行日 平成17年3月16日(2005.3.16)

(24) 登録日 平成16年12月17日(2004.12.17)

(51) Int.Cl.⁷

F I

H O 4 N 5/93

H O 4 N 5/93

G

G 1 O L 21/04

G 1 1 B 20/10

D

G 1 1 B 20/10

G 1 1 B 20/10

3 2 1 Z

G 1 O L 3/02

A

請求項の数 14 (全 15 頁)

(21) 出願番号 特願2002-158898 (P2002-158898)
 (22) 出願日 平成14年5月31日(2002.5.31)
 (65) 公開番号 特開2004-7140 (P2004-7140A)
 (43) 公開日 平成16年1月8日(2004.1.8)
 審査請求日 平成14年5月31日(2002.5.31)

(73) 特許権者 000003078
 株式会社東芝
 東京都港区芝浦一丁目1番1号
 (74) 代理人 100058479
 弁理士 鈴江 武彦
 (74) 代理人 100084618
 弁理士 村松 貞男
 (74) 代理人 100068814
 弁理士 坪井 淳
 (74) 代理人 100092196
 弁理士 橋本 良郎
 (74) 代理人 100091351
 弁理士 河野 哲
 (74) 代理人 100088683
 弁理士 中村 誠

最終頁に続く

(54) 【発明の名称】 音声再生装置および同装置で用いられる音声再生制御方法

(57) 【特許請求の範囲】

【請求項1】

映像データと同期して再生すべき音声データを再生する音声再生装置において、
 前記映像データを再生する他の装置との間の通信を無線通信によって実行する通信装置と

、
 前記通信装置による前記他の装置との通信を利用して、前記他の装置における映像データの再生位置と前記音声再生装置における音声データの再生位置との間のずれ量を検出する手段と、

再生すべき音声データの中から音量レベルが特定値よりも低い音声データ区間を検出する手段と、

前記検出された再生位置のずれ量に基づき、前記検出された音声データ区間内で前記音声データの再生速度を変更する再生速度制御手段とを具備することを特徴とする音声再生装置。

【請求項2】

前記再生速度制御手段は、前記検出された音声データ区間に属する音声データを間引く手段を含むことを特徴とする請求項1記載の音声再生装置。

【請求項3】

前記再生速度制御手段は、前記検出された音声データ区間に無音データを挿入する手段を含むことを特徴とする請求項1記載の音声再生装置。

【請求項4】

前記再生速度制御手段は、前記検出された音声データ区間に属する音声データの、単位時間あたりに再生すべきデータ量を変更する手段を含むことを特徴とする請求項 1 記載の音声再生装置。

【請求項 5】

前記音量レベルが特定値よりも低い音声データ区間を検出する手段は、再生すべき音声データの中から音量レベルが第 1 の音量レベルとそれよりも低い第 2 の音量レベルとの間の範囲に属する第 1 の音声データ区間を検出する手段と、再生すべき音声データの中から音量レベルが前記第 2 の音量レベルよりも低い第 2 の音声データ区間を検出する手段とを含み、

前記再生速度制御手段は、

前記第 1 および第 2 の音声データ区間毎に規定された音声データの再生速度に関する最大許容変更量と、前記検出された再生位置のずれ量とに基づき、前記第 1 および第 2 の各音声データ区間毎にその音声データ区間に属する音声データの再生速度の変更量を決定する手段を含むことを特徴とする請求項 1 記載の音声再生装置。

【請求項 6】

前記通信装置による前記他の装置との通信を利用して、前記他の装置から送信される前記再生すべき音声データを受信する手段と、

前記受信された音声データを一時的に格納する記憶装置とをさらに具備し、

前記音量レベルが特定値よりも低い音声データ区間を検出する手段は、

前記記憶装置に格納されている音声データを解析して再生すべき音声データの中から音量レベルが前記特定値よりも低い音声データ区間を検出する手段を含むことを特徴とする請求項 1 記載の音声再生装置。

【請求項 7】

前記ずれ量を検出する手段は、前記通信装置を用いて特定の時間間隔で前記他の装置と通信することによって、前記他の装置から映像データの再生位置を取得する手段を含むことを特徴とする請求項 1 記載の音声再生装置。

【請求項 8】

映像データと同期して再生すべき音声データを再生する音声再生装置の再生動作を制御する音声再生制御方法において、

前記映像データを再生する他の装置との間の無線通信によって、前記他の装置における映像データの再生位置と前記音声再生装置における音声データの再生位置との間のずれ量を検出するステップと、

再生すべき音声データの中から音量レベルが特定値よりも低い音声データ区間を検出するステップと、

前記検出された再生位置のずれ量に基づき、前記検出された音声データの区間で前記音声データの再生速度を変更する再生速度制御ステップとを具備することを特徴とする音声再生制御方法。

【請求項 9】

前記再生速度制御ステップは、前記検出された音声データの区間に属する音声データを間引くステップを含むことを特徴とする請求項 8 記載の音声再生制御方法。

【請求項 10】

前記再生速度制御ステップは、前記検出された音声データの区間に無音データを挿入するステップを含むことを特徴とする請求項 8 記載の音声再生制御方法。

【請求項 11】

前記再生速度制御ステップは、前記検出された音声データの区間に属する音声データの、単位時間あたりに再生すべきデータ量を変更するステップを含むことを特徴とする請求項 8 記載の音声再生制御方法。

【請求項 12】

前記音量レベルが特定値よりも低い音声データ区間を検出するステップは、再生すべき音声データの中から音量レベルが第 1 の音量レベルとそれよりも低い第 2 の音量レベルとの

10

20

30

40

50

間の範囲に属する第 1 の音声データ区間を検出するステップと、再生すべき音声データの中から音量レベルが前記第 2 の音量レベルよりも低い第 2 の音声データ区間を検出するステップとを含み、

前記再生速度制御ステップは、

前記第 1 および第 2 の音声データ区間毎に規定された音声データの再生速度に関する最大許容変更量と、前記検出された再生位置のずれ量とに基づき、前記第 1 および第 2 の各音声データ区間毎にその音声データ区間に属する音声データの再生速度の変更量を決定するステップを含むことを特徴とする請求項 8 記載の音声再生制御方法。

【請求項 13】

前記他の装置から無線信号によって送信される前記再生すべき音声データを受信するステップと、

前記受信された音声データを記憶装置に一時的に格納するステップとをさらに具備し、

前記音量レベルが特定値よりも低い音声データ区間を検出するステップは、

前記記憶装置に格納されている音声データを解析して再生すべき音声データの中から音量レベルが前記特定値よりも低い音声データ区間を検出するステップを含むことを特徴とする請求項 8 記載の音声再生制御方法。

【請求項 14】

前記ずれ量を検出するステップは、特定の時間間隔で前記他の装置と無線通信することによって、前記他の装置から映像データの再生位置を取得するステップを含むことを特徴とする請求項 8 記載の音声再生制御方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は映像データと同期して再生すべき音声データを再生する音声再生装置および同装置で用いられる音声再生制御方法に関する。

【0002】

【従来の技術】

近年、動画、音声などのマルチメディアデータの再生速度を変更するための技術が開発されている。

【0003】

特開平 9 - 261613 号公報には、マルチメディアデータを受信および再生する装置において再生速度を可変にする再生方法が開示されている。この方法は、受信バッファ内のデータ残量に応じて再生速度を変更する。すなわち、この方法では、受信バッファ内のデータ残量が多い場合には再生速度を上げ、受信バッファ内のデータ残量が少ない場合には再生速度を下げるという制御が実行され、これにより受信バッファのオーバーフロー/アンダーフローの発生を防いでいる。

【0004】

特開平 5 - 135228 号公報には、音声信号の無音期間を拡大させたり縮小させたりすることで、高音質を維持しつつ遅き聴き再生や早聴き再生を実現する技術が開示されている。

【0005】

【発明が解決しようとする課題】

ところで、最近では、互いに同期して再生すべき映像データと音声データとをそれぞれ別個の装置で再生する技術の実現が要求されている。

【0006】

映像と音声のような同期したメディアをそれぞれ異なる装置で再生する場合には、映像の再生タイミング制御のために参照するクロックと音声の再生タイミング制御のために参照するクロックとが異なり、これが映像と音声との同期のずれが生じる原因となる。また同期のずれを解消するために映像データまたは音声データの再生速度を変更すると、視聴者に不自然な印象を与えてしまうことになる。

10

20

30

40

50

【 0 0 0 7 】

本発明は上述の事情を考慮してなされたものであり、視聴者に不自然な印象を与えることなく、他の装置で再生される映像データとの間の同期のずれに応じて音声データの再生速度を変更することが可能な音声再生装置および同装置で用いられる音声再生制御方法を提供することを目的とする。

【 0 0 0 8 】

【課題を解決するための手段】

上述の課題を解決するため、本発明は、映像データと同期して再生すべき音声データを再生する音声再生装置において、前記映像データを再生する他の装置との間の通信を無線信号によって実行する通信装置と、前記通信装置による前記他の装置との通信を利用して、前記他の装置における映像データの再生位置と前記音声再生装置における音声データの再生位置との間のずれ量を検出する手段と、再生すべき音声データの中から音量レベルが特定値よりも低い音声データ区間を検出する手段と、前記検出された再生位置のずれ量に基づき、前記検出された音声データ区間内で前記音声データの再生速度を変更する再生速度制御手段とを具備することを特徴とする。

10

【 0 0 0 9 】

この音声再生装置においては、通信装置による他の装置との無線信号を用いた通信を利用して、他の装置における映像データの再生位置と前記音声再生装置における音声データの再生位置との間のずれ量が検出される。さらに、再生すべき音声データの中から音量レベルが特定値よりも低い音声データ区間が検出される。再生位置のずれ量に基づく音声データの再生速度の変更は、検出された音声データ区間内で実行される。これにより、視聴者に不自然な印象を与えることなく、他の装置で再生される映像データとの間の同期のずれに応じて音声データの再生速度を変更することが可能となる。

20

【 0 0 1 0 】

【発明の実施の形態】

以下、図面を参照して本発明の実施の形態を説明する。

図 1 には、本発明の一実施形態に係る音声再生装置が示されている。この音声再生装置 20 は、映像再生装置 10 による動画などの映像データの再生と同期して音声データを再生するように構成されている。例えば、映像データとそれと同期して再生すべき音声データを含むオーディオ・ビデオコンテンツを再生する場合、映像データの再生は映像再生装置 10 によって行われ、音声データの再生は音声再生装置 20 によって行われる。

30

【 0 0 1 1 】

音声再生装置 20 は、再生すべき音声データを有線または無線の通信路を介して外部機器から受信しながら、その受信した音声データを再生する。音声再生装置 20 への音声データの送信には、映像再生装置 10 から音声再生装置 20 に音声データを送信する方法、または外部のコンテンツサーバから映像データおよび音声データをそれぞれ映像再生装置 10 および音声再生装置 20 に同時に送信する方法が用いられる。以下では、前者の方法を用いる場合を想定することとする。

【 0 0 1 2 】

オーディオ・ビデオコンテンツを再生する場合、映像再生装置 10 は、オーディオ・ビデオコンテンツを映像データと音声データに分離し、音声データを音声再生装置 20 に送信しながら、映像データの再生を行う。音声データの送信には、例えば bluetooth (R)、などに代表される近距離無線通信規格の無線通信が用いられる。この場合、映像再生装置 10 は、音声データをアンテナ 101 から無線信号によって送信し、音声再生装置 20 は映像再生装置 10 から送信される音声データをアンテナ 201 を介して受信する。

40

【 0 0 1 3 】

音声再生装置 20 は映像再生装置 10 から送信される音声データを受信しながら、その受信した音声データの再生を行う。音声データの再生期間中は、音声再生装置 20 は、映像再生装置 10 との間で定期的に通信して、映像再生装置 10 における映像データの再生位

50

置と音声再生装置 20 における音声データの再生位置との間のずれ量（映像と音声の同期のずれ量）を検出する。音声再生装置 20 は、映像と音声の同期再生を実現するために、検出したずれ量に応じて音声データの再生速度を変更する。ここで、音声データの再生速度の変更は、音声データの音量レベルが予め決められた特定値よりも低い音声データ区間で実行される。

【0014】

次に、図 3 および図 4 を参照して、映像再生装置 10 および音声再生装置 20 それぞれの構成を説明する。

【0015】

図 3 には、映像再生装置 10 のシステム構成が示されている。映像再生装置 10 は、例えば PDA (Personal Data Assistants)、ポータブルパーソナルコンピュータ、移動電話機、のような携帯型電子機器として実現されている。この映像再生装置 10 は、図示のように、CPU 11、時刻管理部 12、記憶装置 13、無線通信デバイス 14、映像/音声分離部 15、復号部 16、映像再生部 17、およびディスプレイ 18 を備えている。

10

【0016】

CPU 11 は映像再生装置 10 の動作を制御するために設けられたプロセッサである。この CPU 11 はバス 100 を介して映像再生装置 10 の他のコンポーネントと接続されており、それら各コンポーネントと連携して動作する。時刻管理部 12 は時刻を計時する時計デバイスから構成されており、映像データの再生動作の基準となる時刻情報を提供する

20

【0017】

記憶装置 13 は例えば不揮発性メモリデバイスから構成されており、ここには、再生対象のオーディオ・ビデオコンテンツが格納される。このオーディオ・ビデオコンテンツは、符号化された映像データのストリームとそれと同期して再生すべき符号化された音声データのストリームとから構成されている。これら映像データストリームと音声データストリームは多重化されている。オーディオ・ビデオコンテンツのデータフォーマットとしては、例えば AVI 形式 (Audio Video Interleave Format) を使用することが出来る。AVI 形式のコンテンツデータにおいては、映像および音声のメディア毎にそれに対応するサンプリングレートが記録されている。

30

【0018】

映像/音声分離部 15 は再生対象のコンテンツデータを映像データと音声データに分離する。コンテンツデータから分離された映像データは復号部 16 に送られ、またコンテンツデータから分離された音声データは無線通信デバイス 14 に送られる。

【0019】

無線通信デバイス 14 は音声再生装置 20 との間の無線接続を確立して、音声再生装置 20 との間の無線通信を実行する。無線通信は、例えば Bluetooth (R) の規格に準拠した手順で実行される。無線通信デバイス 14 は、CPU 11 の制御に従い、符号化された音声データを無線信号によって音声再生装置 20 に送信する。

【0020】

40

復号部 16 は符号化された映像データを復号して、圧縮形式の映像データを表示可能な形式に変換する。映像再生部 17 はディスプレイ 18 を制御し、復号部 16 によって復号された映像データをディスプレイ 18 上に表示するための再生動作を実行する。この場合、映像データの再生速度は、時刻管理部 12 によって管理される時刻情報と映像データのサンプリングレートとに従って制御される。

【0021】

映像再生部 17 による映像データの再生は、例えば音声データの無線送信を開始してから一定時間経過した時点で開始される。この時間の値は、例えば、音声データの無線送信を開始してから、音声再生装置 20 がその音声データの再生が可能になるまでに要する時間に基づいて予め決定されている。これにより、音声再生装置 20 による音声データの再生

50

と映像再生部 17 による映像データの再生とを同時に開始させることが可能となる。

【0022】

図 4 には、音声再生装置 20 のシステム構成が示されている。音声再生装置 20 は、上述したようにワイヤレスヘッドフォンとして実現されている。この音声再生装置 20 は、図示のように、CPU 21、時刻管理部 22、無線通信デバイス 23、復号部 24、バッファメモリ 25、音声再生部 26、音声出力装置 27、映像再生位置取得部 28、同期ずれ量検出部 29、音出力レベル検出部 30、および音声再生速度制御部 31 を備えている。

【0023】

CPU 21 は音声再生装置 20 の動作を制御するために設けられたプロセッサである。この CPU 21 はバス 200 を介して音声再生装置 20 の他のコンポーネントと接続されており、それら各コンポーネントと連携して動作する。時刻管理部 22 は時刻を計時する時計デバイスから構成されており、音声データの再生動作の基準となる時刻情報を提供する。

10

【0024】

無線通信デバイス 23 は映像再生装置 10 との間の無線接続を確立して、映像再生装置 10 との間の無線通信を実行する。無線通信は、例えば Bluetooth (R) の規格に準拠した手順で実行される。無線通信デバイス 23 は、CPU 21 の制御に従い、映像再生装置 10 から送信される符号化された音声データを受信する。

【0025】

復号部 24 は無線通信デバイス 23 によって受信された符号化された音声データを復号して、圧縮形式の音声データを再生可能な形式に変換する。復号された音声データはバッファメモリ 25 に格納される。このバッファメモリ 25 は、再生すべき音声データを一時的に保持するための記憶装置である。このバッファメモリ 25 には、例えば、50 msec ~ 600 msec 程度の再生時間に相当する音声データが蓄積される。

20

【0026】

音声再生部 26 は、バッファメモリ 25 に蓄積されている音声データをスピーカ又はヘッドフォンからなる音声出力部 27 を通じて音として再生出力する。この音声再生動作においては、バッファメモリ 25 に蓄積されている音声データをデジタル形式からアナログ形式に変換する処理などが行われる。音声データの再生速度は、基本的には、時刻管理部 22 によって管理される時刻情報と音声データのサンプリングレートとに従って決定されるが、音声再生速度制御部 31 の制御によって変更することが出来る。

30

【0027】

映像再生装置 10 における映像再生タイミングの制御は音声再生装置 20 で行われる音声再生タイミングの制御とは異なるクロック (時刻情報) が参照されているため、音声と映像の同期にずれが生じることが考えられる。もし映像の再生タイミング制御と音声の再生タイミング制御とで同一のクロックを参照することが出来れば、音声と映像の同期再生を容易に実現できる。しかし、本実施形態のように映像と音声を再生する装置が互いに異なる場合には、一方の装置から他方の装置のクロックを頻繁に参照することは實際上困難である。

【0028】

40

そこで、本実施形態では、音声の再生タイミング制御は、上述したように、映像の再生タイミング制御とは別のクロック (時刻管理部 22 の時刻情報) を参照しながら実行する。音声の再生期間中は、音声再生装置 20 は、映像再生装置 10 との間の無線通信を利用して、定期的に映像再生装置 10 から映像の再生位置を取得する。そして、音声再生装置 20 は、取得した映像再生位置を用いて音声再生位置の修正を行うことで同期のずれを解消する。

【0029】

このような音声再生位置の修正処理は、映像再生位置取得部 28、同期ずれ量検出部 29、音出力レベル検出部 30、および音声再生速度制御部 31 によって実行される。

【0030】

50

映像再生位置取得部 28 および同期ずれ量検出部 29 は、無線通信デバイス 23 による映像再生装置 10 との通信を利用して、映像再生装置 10 における映像データの再生位置と音声再生装置 20 における音声データの再生位置との間のずれ量を検出する。

【0031】

すなわち、映像再生位置取得部 28 は、無線通信デバイス 23 を用いて映像再生位置の取得要求を映像再生装置 10 に送信し、映像再生装置 10 から現在の映像データの再生位置を取得する。映像再生位置の取得要求は、音声データの再生期間中に定期的に映像再生装置 10 に送信される。同期ずれ量検出部 29 は、音声再生部 26 による現在の音声データの再生位置と、映像再生位置取得部 28 によって取得された現在の映像データの再生位置とに基づいて、映像データに対する音声データの再生位置のずれ量（同期ずれ量）を検出する。現在の音声データの再生位置は、例えば、音声データの再生開始時点から現在までの経過時間で与えられ、同様に、現在の映像データの再生位置も、映像データの再生開始時点から現在までの経過時間で与えられる。

10

【0032】

映像再生位置取得部 28 および同期ずれ量検出部 29 による同期ずれ量の検出処理は、例えば、NTP (Network Time Protocol) によるクロック同期を利用して実行することができる。この NTP は、ネットワークを介して接続される装置間の時刻を合わせるプロトコルである。すなわち、映像再生装置 10 および音声再生装置 20 はそれぞれ自身の時刻管理部 12, 22 を参照して映像及び音声の再生タイミングを制御しているので、音声再生装置 20 の時刻管理部 22 が示す時刻と映像再生装置 10 の時刻管理部 12 が示す時刻との差を NTP を用いて調べることにより、映像データの再生位置と音声データの再生位置との間のずれ量を検出することができる。

20

【0033】

音出力レベル検出部 30 は、バッファメモリ 25 に蓄積されている音声データを解析することによって、これから再生される音声データのストリームがどのような音量レベル（音出力レベル）であるかを検出する。この音量レベル検出処理によって、これから再生される音声データのストリームの中から音量レベルが予め決められた特定値よりも低い音声データ区間が検出される。

【0034】

音声再生速度制御部 31 は、同期のずれを解消するために、音声再生部 26 による音声データの再生速度を可変制御するように構成されている。音声再生速度の変更は、これから再生される音声データのストリームの中から検出された低音量レベルの区間内で実行される。音出力レベル検出部 30 によって検出された音声データの音量レベルが大きい音声データ区間で音声再生速度の大きな変更を行うと視聴者に不自然な印象を与える可能性が高く、一方音量レベルが小さい区間では音声再生速度の変更を行っても視聴者に不自然な印象を与える可能性が低いからである。

30

【0035】

音声再生速度制御部 31 は、図 5 に示すように、音声再生速度決定部 311 と、音声再生速度変更部 312 とから構成されている。音声再生速度決定部 311 は、同期ずれ量検出部 29 によって検出された同期ずれ量と、音出力レベル検出部 30 による音量レベルの検出結果とに基づいて、音声再生速度の変更量を決定する。音声再生速度の変更量は、なるべく音量レベルが小さい区間で音声再生速度の変更が行われるように、あらかじめ定められた規則に従って決定される。この規則は、音量レベル毎に音声再生速度の最大許容変更量を規定したものである。

40

【0036】

音声再生速度変更部 312 は、音声再生速度決定部 311 によって決定された音声再生速度の変更量に従い、音声再生部 26 による音声データの再生速度を変更する。音声再生速度の変更は、1) 単位時間あたりに再生すべき音声データの量の増減、2) 音声データの間引き、3) 音声データ（無音データ）の挿入、などによって行われる。

【0037】

50

例えば、映像データの再生位置が音声データの再生位置よりも進んでいる場合には、単位時間あたりに再生すべき音声データの量の増加、あるいは音声データの間引き（削除を含む）を行うことで、同期のずれを解消する。一方、映像データの再生位置が音声データの再生位置よりも遅れている場合には、単位時間あたりに再生すべき音声データの量の減少、あるいは音声データ（無音データ）の挿入を行うことで、同期のずれを解消する。

【0038】

次に、音量レベル検出処理の例について説明する。

図6は、バッファメモリ25に蓄積される音声データを模式的に示している。音声データの音量レベルは、時間的に連続するサンプルデータ（図6の黒丸）を結ぶデジタル音声信号の波形の振幅の大きさによって与えられる。各サンプルデータは、アナログ音声信号を、例えば44.1kHzのサンプリングレートでサンプリングすることによって得られたものである。

10

【0039】

音出力レベル検出部30は、例えば、音声データの波形の振幅の大きさの変化を調べることにより、バッファメモリ25に蓄積される音声データのストリームの中から、“音量レベル＝大”の音声区間と、“音量レベル＝中”の音声区間と、“音量レベル＝小”の音声区間とを検出する。音声再生速度の変更は“音量レベル＝中”の音声区間または“音量レベル＝小”の音声区間で行われ、“音量レベル＝大”の音声区間では行われない。“音量レベル＝中”の音声区間における音声再生速度の最大許容変更量は、“音量レベル＝小”の音声区間における音声再生速度の最大許容変更量よりも少ない。

20

【0040】

音出力レベルの検出には、第1の閾値とそれよりも小さい第2の閾値とが用いられる。振幅の大きさつまり音量レベルが、第2の閾値よりも低い音声データの区間は“音量レベル＝小”の区間として検出され、振幅の大きさつまり音量レベルが、第1の閾値と第2の閾値との間に属する音声データの区間は“音量レベル＝中”の区間として検出される。第1および第2の閾値の値はそれぞれ予め規定された固定値である。

【0041】

次に、本実施形態の音声再生装置20によって実行される音声再生制御処理の動作について説明する。

【0042】

30

図7のフローチャートには、音声再生装置20が映像と音声の同期のずれを検出し、その検出結果に基づいて音声再生速度を変更する手順が示されている。

【0043】

音声再生装置20は、動画再生装置10から送信される音声データを受信しながら、その音声データを再生する。音声データの再生期間中は、音声再生装置20は、動画再生装置10における映像再生と同期して音声を再生するために、あらかじめ定められた時間間隔毎に、同期を取るための処理を実行する。

【0044】

すなわち、音声再生装置20は、最後に同期をとる処理が行われてからの経過時間から、同期をとる必要があるか否かを判断する（ステップS101）。最後に同期をとる処理が行われてからの経過時間があらかじめ定められた時間以上であれば（ステップS101のYES）、音声再生装置20は、以下の同期処理を行う。

40

【0045】

まず、音声再生装置20の映像再生位置取得部28は、無線通信デバイス23による映像再生装置10との通信を利用して、映像再生装置10にて現在再生中の映像の再生位置を取得する（ステップS102）。次いで、同期ずれ量検出部29は、ステップS102にて取得された映像再生装置10における映像再生位置と音声再生装置20における音声再生位置とを比較し、動画と音声の再生位置間にずれがあるか否かを判断する（ステップS103）。

【0046】

50

動画と音声の再生位置間にずれがあると判断された場合には（ステップS103のYES）、音声再生装置20の音出力レベル検出部30は、現在再生しているの音声の再生位置以降の、バッファメモリ25に蓄えられている音声データについて音量レベル検出を行う（ステップS104）。この後、音声再生速度制御部31の音声再生速度決定部311は、ステップS103において検出された映像と音声間の再生位置のずれ量と、ステップS104において検出された音声データの音量レベルとに基づいて、あらかじめ定められた規則に従って再生速度変更を行うべき音声データ区間とその区間における再生速度の変更量を決定する（ステップS105）。

【0047】

音声再生速度の変更量を決定する際に用いられる規則の例を図8に示す。図8において、1行目はステップS104により検出された音声データの音量レベルに対応している。2行目は1行目の音量レベルそれぞれに対応する、音声再生速度の最大許容変更量、つまり単位時間あたりに再生すべき音声データ量の最大許容増減幅を示している。同様に3行目は1行目の音量それぞれに対応する、音声データの挿入、間引き（削除）を行う際の最大許容時間、つまり単位時間あたりに挿入、削除が可能な音声データの最大許容データ長に対応する時間、を示している。

【0048】

音声再生速度の最大許容変更量、および音声データの挿入、削除の最大許容時間は、いずれも音量レベルが小さいほど大きな値となり、音量レベルが大きいほど小さい値となるように設定される。

【0049】

音声再生速度変更、または音声データの挿入、削除は、映像と音声のずれが解消されるまで複数回続けられる。またあらかじめ定められた時間のうちに同期のずれが解消されない場合には、直ちに同期のずれを解消する時間だけ音声データの削除、挿入を行う。

【0050】

音声再生装置11において、動画と音声の同期のずれを解消する際には、映像の再生位置を基準に音声の再生位置を修正する。映像の再生位置が音声の再生位置よりも進んでいる場合には、単位時間あたりに再生すべき音声データの量の増加、あるいは音声データの削除を行うことで、同期のずれを解消する。一方、映像データの再生位置が音声データの再生位置よりも遅れている場合には、単位時間あたりに再生すべき音声データの量の減少、あるいは音声データ（無音データ）の挿入を行うことで、同期のずれを解消する。

【0051】

このようにして、ステップS105では単位時間あたりに再生すべき音声データの増減量、または音声データの挿入、削除の時間が決定される。この決定に従い、音声再生速度制御部31の音声再生速度変更部312は音声再生部26を制御し、音声再生部26の音声再生速度を変更する（ステップS106）。

【0052】

ここで、ステップS104～S106の処理の具体例について、図9を参照しながら説明する。

【0053】

図9は音声データの再生速度が変更される様子を示している。図9に示されている3つの音声データの波形の内、上部の音声データの波形は、同期ずれが生じていない正常な状態を示している。また中央の音声データの波形は、同期ずれが生じ、音声データの再生が映像の再生よりも遅れている状態を示している。下部の音声データの波形は、音声再生速度の変更処理によって同期ずれを解消した状態を示している。

【0054】

いま、図中に示したAの時点で、100msecの遅れを検出した場合を想定する。この場合、まず、バッファメモリ25上に蓄えられたAの位置以降における音声データの音量レベルが検出される（ステップS104）。そして検出された音量レベルに基づき、図8の規則に従い音声再生速度の変更量が決定される（ステップS105）。

10

20

30

40

50

【0055】

AからBまでの区間(100 msec)においては検出された音量レベルは“大”であるので、規則から、AからBまでの区間における音声再生速度の変更量は、0 msecとなる。しかし、B以降Dまでの区間(600 msec)における音量レベルは“小”であるので、規則から、一秒あたりにつき±500 msecの音声データの挿入又は間引きを行うことが許容される。検出された遅れは100 msecである。この遅れは、B以降Dまでの区間(600 msec)内の音声データを100 msec分だけ間引くことにより解消される。すなわち、B以降Dまでの区間(600 msec)内の音声データは500 msecに短縮され、これによってCの時点では音声データの再生位置は映像データの再生位置と一致する。

10

【0056】

なお、NTPを用いて同期ずれを検出する場合には、再生位置のずれが解消されるCの時点で時刻管理部22の時刻も100 msec分だけ進めることが好ましい。

【0057】

以上詳述したように本実施形態によれば、映像再生装置10と音声再生装置2との間で厳密な時刻合わせを行わずとも、音声再生速度変更処理によって同期ずれ量にある一定の範囲内に抑えることが可能となる。よって、映像と音声をそれぞれ異なる装置で再生するシステムにおいても、いわゆるリップシンク(lip sync)を容易に実現することが出来る。特に、音声再生速度の変更は音量レベルが低い音声データ区間内で行っている

20

【0058】

なお、必ずしも映像再生装置10が音声再生装置20に音声データを送信する必要はなく、例えば、図10に示すように、コンテンツサーバ30が、映像データおよび音声データをアンテナ301からそれぞれ無線信号によって、映像再生装置10および音声再生装置20に送信するようにしても良い。この場合でも、音声再生装置20は、映像再生装置10から定期的に映像再生位置を取得することにより、映像と音声の同期ずれ量を検出することが出来る。

【0059】

また、本実施形態の音声再生装置20における音声再生制御処理は、コンピュータ上で実行されるプログラムによって実現することも出来る。この場合、上述の音声再生制御処理の手順を含むコンピュータプログラムを記憶したコンピュータ読み取り可能な記憶媒体を用意し、そのコンピュータプログラムを、通信機能を持つ通常のコンピュータに導入して実行するだけで、本実施形態と同様の効果を容易に得ることが可能となる。

30

【0060】

また、本発明は、上記実施形態に限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。更に、上記実施形態には種々の段階の発明が含まれており、開示される複数の構成要件における適宜な組み合わせにより種々の発明が抽出され得る。例えば、実施形態に示される全構成要件から幾つかの構成要件が削除されても、発明が解決しようとする課題の欄で述べた課題が解決でき、発明の効果の欄で述べられている効果が得られる場合には、この構成要件が削除された構成が発明として抽出され得る。

40

【0061】

【発明の効果】

以上説明したように、本発明によれば、視聴者に不自然な印象を与えることなく、他の装置で再生される映像データとの間の同期のずれに応じて音声データの再生速度を変更することが可能となる。

【図面の簡単な説明】

【図1】本発明の一実施形態に係る音声再生装置を示す図。

【図2】図1に示した音声再生装置の具体的の実現形態の一例であるワイヤレスヘッドフォンを示す図。

50

【図 3】図 1 に示した音声再生装置によって再生される音声データと同期して再生すべき映像データを再生する映像再生装置の構成の例を示すブロック図。

【図 4】図 1 に示した音声再生装置の構成の例を示すブロック図。

【図 5】図 1 に示した音声再生装置に設けられた音声再生速度制御部の構成を示すブロック図。

【図 6】図 1 に示した音声再生装置に設けられた音出力レベル検出部が実行する音量レベル検出処理を説明するための図。

【図 7】図 1 に示した音声再生装置によって実行される音声再生制御処理の手順を示すフローチャート。

【図 8】図 1 に示した音声再生装置が音声再生速度の変更量を決定するために使用する規則の例を説明するための図。 10

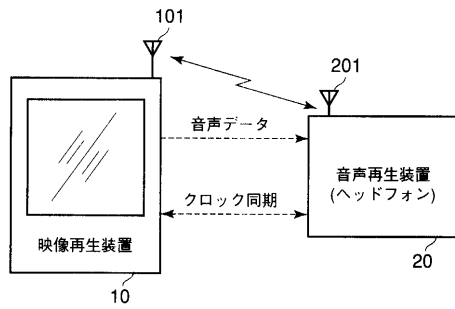
【図 9】図 1 に示した音声再生装置において音声データの再生速度が変更される様子を示す図。

【図 10】図 1 に示した音声再生装置に音声データをサーバから送信するシステムの例を示すブロック図。

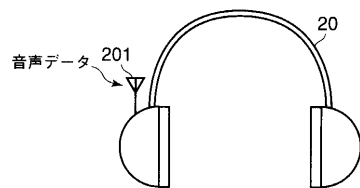
【符号の説明】

- 1 0 ... 映像再生装置
- 2 0 ... 音声再生装置
- 2 1 ... C P U
- 2 2 ... 時刻管理部
- 2 3 ... 無線通信デバイス
- 2 4 ... 復号部
- 2 5 ... バッファメモリ
- 2 6 ... 音声再生部
- 2 7 ... 音声出力装置
- 2 8 ... 映像再生位置取得部
- 2 9 ... 同期ずれ量検出部
- 3 0 ... 音出力レベル（音量レベル）検出部
- 3 1 ... 音声再生速度制御部

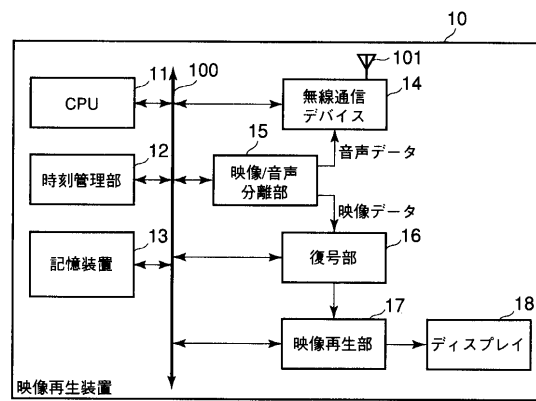
【図 1】



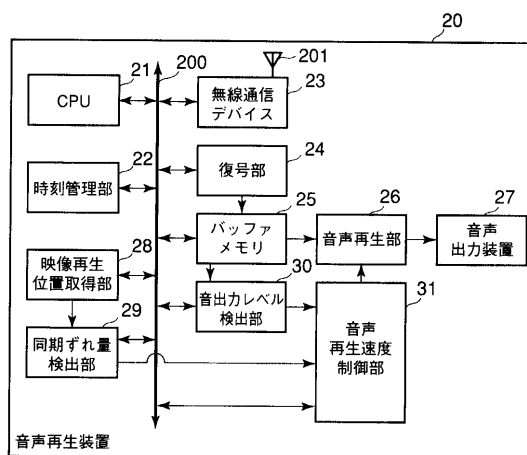
【図 2】



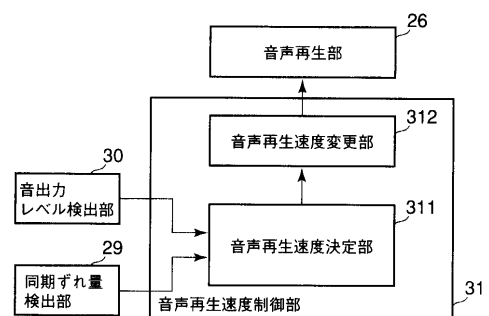
【図 3】



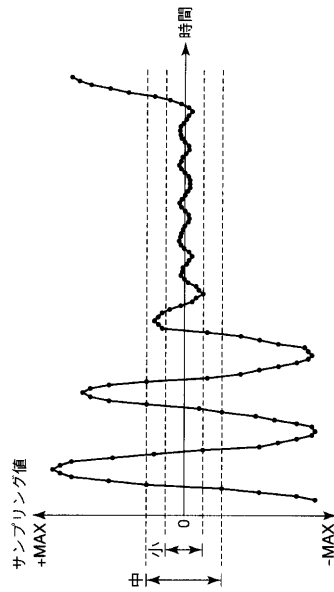
【図 4】



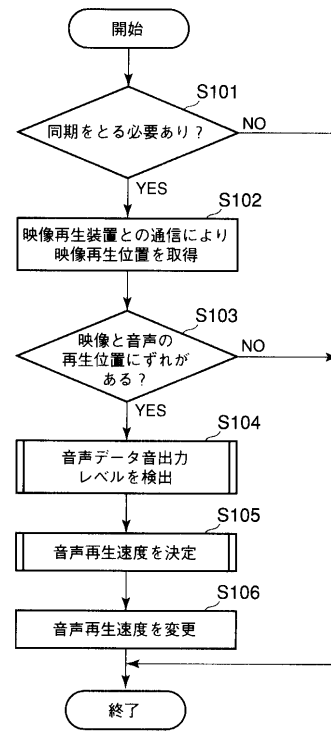
【図 5】



【図 6】



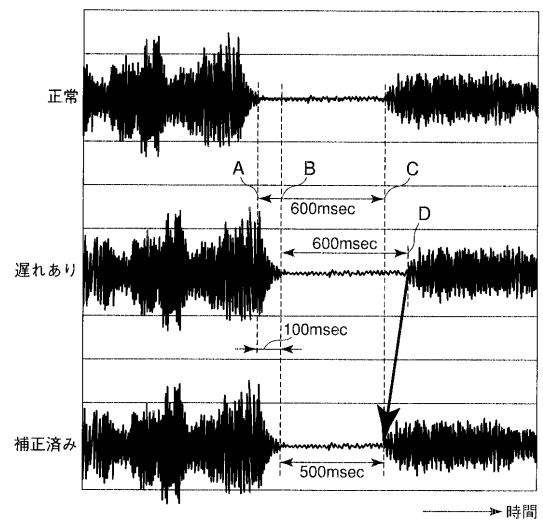
【図 7】



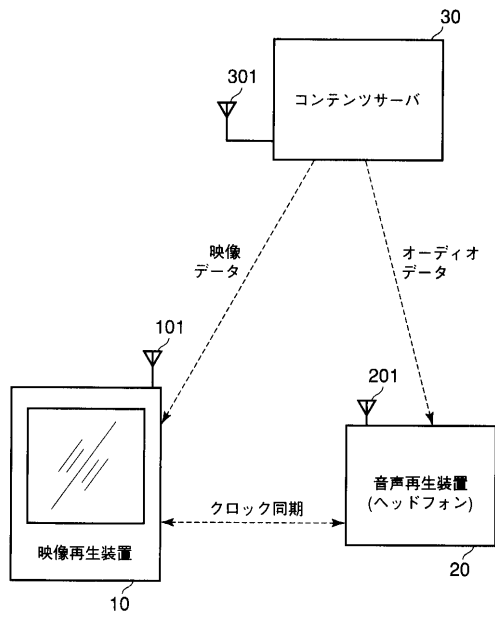
【図 8】

検出された 音出力レベル	小	中	大
音声再生速度の 最大許容変更量 (一秒あたり)	±1.0% (±8msec)	±0.5% (±4msec)	±0.0%
音声データの挿入、 削除の最大許容時間 (一秒あたり)	±500msec	±250msec	±0msec

【図 9】



【図 10】



フロントページの続き

(74)代理人 100070437

弁理士 河井 将次

(72)発明者 牛丸 主税

東京都青梅市末広町 2 丁目 9 番地 株式会社東芝青梅工場内

審査官 鈴木 明

(56)参考文献 特開平 0 9 - 0 6 5 3 0 3 (J P , A)

(58)調査した分野(Int.Cl.⁷, D B 名)

H04N 5/76-5/956

G11B 20/10-20/12

G10L 21/04