

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5074834号
(P5074834)

(45) 発行日 平成24年11月14日(2012.11.14)

(24) 登録日 平成24年8月31日(2012.8.31)

(51) Int.Cl.		F I	
HO4N	7/14	(2006.01)	HO4N 7/14
HO4L	12/56	(2006.01)	HO4L 12/56 230B
HO4M	11/00	(2006.01)	HO4M 11/00 302

請求項の数 3 (全 11 頁)

<p>(21) 出願番号 特願2007-172297 (P2007-172297)</p> <p>(22) 出願日 平成19年6月29日 (2007.6.29)</p> <p>(65) 公開番号 特開2009-10863 (P2009-10863A)</p> <p>(43) 公開日 平成21年1月15日 (2009.1.15)</p> <p>審査請求日 平成22年3月16日 (2010.3.16)</p>	<p>(73) 特許権者 000000295 沖電気工業株式会社 東京都港区虎ノ門一丁目7番12号</p> <p>(73) 特許権者 591051645 株式会社OKIソフトウェア 埼玉県蕨市中央一丁目16番8号</p> <p>(74) 代理人 100090620 弁理士 工藤 宣幸</p> <p>(72) 発明者 佐藤 達也 埼玉県蕨市中央1-16-8 沖通信システム株式会社内</p> <p>(72) 発明者 久保木 浩 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社内</p>
--	---

最終頁に続く

(54) 【発明の名称】 音声・映像同期方法、音声・映像同期システム及び音声・映像受信端末

(57) 【特許請求の範囲】

【請求項1】

音声・映像送信端末が音声・映像受信端末への音声信号及び映像信号を並行的にネットワークに送信し、上記音声・映像受信端末が、上記音声・映像送信端末からの音声信号及び映像信号を同期化させる音声・映像同期システムにおいて、

上記音声・映像送信端末は、

自己が有するローカル時計の時刻を含む音声についての音声ローカル時刻通知信号を上記音声・映像受信端末に向けて送信する音声ローカル時刻通知信号送信手段と、

ネットワークで調停済みの自己が有する調停済時計の時刻を含む音声についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信する音声調停済時刻通知信号送信手段と、

自己が有するローカル時計の時刻を含む映像についてのローカル時刻通知信号を上記音声・映像受信端末に向けて送信する映像ローカル時刻通知信号送信手段と、

ネットワークで調停済みの自己が有する調停済時計の時刻を含む映像についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信する映像調停済時刻通知信号送信手段とを備え、

上記音声・映像受信端末は、

上記音声調停済時刻通知信号が与えられたときに、ネットワークで調停済みの自己が有する調停済時計での受信時刻と、その通知信号での時刻と、上記音声ローカル時刻通知信号の当該音声・映像受信端末での再生タイミングに基づいて得た音声についての処理待

ち時間とから、映像についてのオフセット時間を得る映像オフセット時間取得手段と、

上記映像調停済時刻通知信号が与えられたときに、自己が有する上記調停済時計での受信時刻を上記オフセット時間分だけ修正すると共に、上記映像ローカル時刻通知信号が与えられたときに、自己が有するローカル時計の時刻を、その通知信号での時刻に修正する映像時刻修正手段と、

上記映像時刻修正手段が修正した2種類の時刻に基づいて、上記音声・映像送信端末が送信した映像信号を、上記音声・映像送信端末が送信した音声信号に同期化させる処理を行う映像処理手段とを備える

ことを特徴とする音声・映像同期システム。

【請求項2】

音声・映像送信端末が音声・映像受信端末への音声信号及び映像信号を並行的にネットワークに送信し、上記音声・映像受信端末が、上記音声・映像送信端末からの音声信号及び映像信号を同期化させる音声・映像同期方法において、

上記音声・映像送信端末は、音声ローカル時刻通知信号送信手段、音声調停済時刻通知信号送信手段、映像ローカル時刻通知信号送信手段及び映像調停済時刻通知信号送信手段を備え、

上記音声ローカル時刻通知信号送信手段は、自己が有するローカル時計の時刻を含む音声についての音声ローカル時刻通知信号を上記音声・映像受信端末に向けて送信し、

上記音声調停済時刻通知信号送信手段は、ネットワークで調停済みの自己が有する調停済時計の時刻を含む音声についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信し、

上記映像ローカル時刻通知信号送信手段は、自己が有するローカル時計の時刻を含む映像についてのローカル時刻通知信号を上記音声・映像受信端末に向けて送信し、

上記映像調停済時刻通知信号送信手段は、ネットワークで調停済みの自己が有する調停済時計の時刻を含む映像についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信すると共に、

上記音声・映像受信端末は、映像オフセット時間取得手段、映像時刻修正手段及び映像処理手段を備え、

上記映像オフセット時間取得手段は、上記音声調停済時刻通知信号が与えられたときに、ネットワークで調停済みの自己が有する調停済時計での受信時刻と、その通知信号での時刻と、上記音声ローカル時刻通知信号の当該音声・映像受信端末での再生タイミングに基づいて得た音声についての処理待ち時間とから、映像についてのオフセット時間を得、

上記映像時刻修正手段は、上記映像調停済時刻通知信号が与えられたときに、自己が有する上記調停済時計での受信時刻を上記オフセット時間分だけ修正すると共に、上記映像ローカル時刻通知信号が与えられたときに、自己が有するローカル時計の時刻を、その通知信号での時刻に修正し、

上記映像処理手段は、上記映像時刻修正手段が修正した2種類の時刻に基づいて、上記音声・映像送信端末が送信した映像信号を上記音声・映像送信端末が送信した音声信号に同期化させる処理を行う

ことを特徴とする音声・映像同期方法。

【請求項3】

音声・映像送信端末が並行的にネットワークに送信した音声信号及び映像信号が与えられ、上記音声・映像送信端末からの音声信号及び映像信号を同期化させる音声・映像受信端末において、

対向する上記音声・映像送信端末が、

自己が有するローカル時計の時刻を含む音声についての音声ローカル時刻通知信号を上記音声・映像受信端末に向けて送信する音声ローカル時刻通知信号送信手段と、

ネットワークで調停済みの自己が有する調停済時計の時刻を含む音声についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信する音声調停済時刻通知信号送信手段と、

10

20

30

40

50

自己が有するローカル時計の時刻を含む映像についてのローカル時刻通知信号を上記音声・映像受信端末に向けて送信する映像ローカル時刻通知信号送信手段と、

ネットワークで調停済みの自己が有する調停済時計の時刻を含む映像についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信する映像調停済時刻通知信号送信手段とを備えたものであり、

自端末は、

上記音声調停済時刻通知信号が与えられたときに、ネットワークで調停済みの自己が有する調停済時計での受信時刻と、その通知信号での時刻と、上記音声ローカル時刻通知信号の当該音声・映像受信端末での再生タイミングに基づいて得た音声についての処理待ち時間とから、映像についてのオフセット時間を得る映像オフセット時間取得手段と、

上記映像調停済時刻通知信号が与えられたときに、自己が有する上記調停済時計での受信時刻を上記オフセット時間分だけ修正すると共に、上記映像ローカル時刻通知信号が与えられたときに、自己が有するローカル時計の時刻を、その通知信号での時刻に修正する映像時刻修正手段と、

上記映像時刻修正手段が修正した２種類の時刻に基づいて、上記音声・映像送信端末が送信した映像信号を上記音声・映像送信端末が送信した音声信号に同期化させる処理を行う映像処理手段とを備える

ことを特徴とする音声・映像受信端末。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は音声・映像同期方法、音声・映像同期システム及び音声・映像受信端末に関し、例えば、テレビ電話システムやテレビ会議システムに適用し得るものである。

【背景技術】

【０００２】

音声と映像の同期をとる転送制御方法として、特許文献１に記載されたものがある。特許文献１に記載の方法は、ＡＶ機能を持った端末において、送信する側は、取り込んだ音声と映像を別々にパケット化すると共に、一定の間隔でユニークに識別可能なマークを、音声と映像のそれぞれのパケットに同じように付与して送信するものであった。

【特許文献１】特開平７－５０８１８号公報

【発明の開示】

【発明が解決しようとする課題】

【０００３】

しかしながら、特許文献１に記載の方法は、送信側にて、音声及び映像パケットへ識別可能なマークを付与し、受信側にて、それら識別可能なマークを認識することにより、同期を実現するものであり、送信側及び受信側の双方に同期のための特殊な構成、機能を盛り込む必要がある。ここで、このような機能をソフトウェアで実現する場合には、結果として、特殊な機能を盛り込むことにより、送信側及び受信側の双方で、ソフトウェアの大きな処理負荷となる可能性もある。

【０００４】

そのため、受信側における音声及び映像の同期確立を簡易に実現できる音声・映像同期方法、音声・映像同期システム及び音声・映像受信端末が望まれている。

【課題を解決するための手段】

【０００５】

第１の本発明は、音声・映像送信端末が音声・映像受信端末への音声信号及び映像信号を並行的にネットワークに送信し、上記音声・映像受信端末が、上記音声・映像送信端末からの音声信号及び映像信号を同期化させる音声・映像同期システムにおいて、（１）上記音声・映像送信端末は、（１－１）自己が有するローカル時計の時刻を含む音声についての音声ローカル時刻通知信号を上記音声・映像受信端末に向けて送信する音声ローカル時刻通知信号送信手段と、（１－２）ネットワークで調停済みの自己が有する調停済時計

10

20

30

40

50

の時刻を含む音声についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信する音声調停済時刻通知信号送信手段と、(1-3)自己が有するローカル時計の時刻を含む映像についてのローカル時刻通知信号を上記音声・映像受信端末に向けて送信する映像ローカル時刻通知信号送信手段と、(1-4)ネットワークで調停済みの自己が有する調停済時計の時刻を含む映像についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信する映像調停済時刻通知信号送信手段とを備え、(2)上記音声・映像受信端末は、(2-1)上記音声調停済時刻通知信号が与えられたときに、ネットワークで調停済みの自己が有する調停済時計での受信時刻と、その通知信号での時刻と、上記音声ローカル時刻通知信号の当該音声・映像受信端末での再生タイミングに基づいて得た音声についての処理待ち時間とから、映像についてのオフセット時間を得る映像オフセット時間取得手段と、(2-2)上記映像調停済時刻通知信号が与えられたときに、自己が有する上記調停済時計での受信時刻を上記オフセット時間分だけ修正すると共に、上記映像ローカル時刻通知信号が与えられたときに、自己が有するローカル時計の時刻を、その通知信号での時刻に修正する映像時刻修正手段と、(2-3)記映像時刻修正手段が修正した2種類の時刻に基づいて、上記音声・映像送信端末が送信した映像信号を上記音声・映像送信端末が送信した音声信号に同期化させる処理を行う映像処理手段とを備えることを特徴とする。

10

【0006】

第2の本発明は、音声・映像送信端末が音声・映像受信端末への音声信号及び映像信号を並行的にネットワークに送信し、上記音声・映像受信端末が、上記音声・映像送信端末からの音声信号及び映像信号を同期化させる音声・映像同期方法において、(1)上記音声・映像送信端末は、音声ローカル時刻通知信号送信手段、音声調停済時刻通知信号送信手段、映像ローカル時刻通知信号送信手段及び映像調停済時刻通知信号送信手段を備え、(1-1)上記音声ローカル時刻通知信号送信手段は、自己が有するローカル時計の時刻を含む音声についての音声ローカル時刻通知信号を上記音声・映像受信端末に向けて送信し、(1-2)上記音声調停済時刻通知信号送信手段は、ネットワークで調停済みの自己が有する調停済時計の時刻を含む音声についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信し、(1-3)上記映像ローカル時刻通知信号送信手段は、自己が有するローカル時計の時刻を含む映像についてのローカル時刻通知信号を上記音声・映像受信端末に向けて送信し、(1-4)上記映像調停済時刻通知信号送信手段は、ネットワークで調停済みの自己が有する調停済時計の時刻を含む映像についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信すると共に、(2)上記音声・映像受信端末は、映像オフセット時間取得手段、映像時刻修正手段及び映像処理手段を備え、(2-1)上記映像オフセット時間取得手段は、上記音声調停済時刻通知信号が与えられたときに、ネットワークで調停済みの自己が有する調停済時計での受信時刻と、その通知信号での時刻と、上記音声ローカル時刻通知信号の当該音声・映像受信端末での再生タイミングに基づいて得た音声についての処理待ち時間とから、映像についてのオフセット時間を得、(2-2)上記映像時刻修正手段は、上記映像調停済時刻通知信号が与えられたときに、自己が有する上記調停済時計での受信時刻を上記オフセット時間分だけ修正すると共に、上記映像ローカル時刻通知信号が与えられたときに、自己が有するローカル時計の時刻を、その通知信号での時刻に修正し、(2-3)上記映像処理手段は、上記映像時刻修正手段が修正した2種類の時刻に基づいて、上記音声・映像送信端末が送信した映像信号を上記音声・映像送信端末が送信した音声信号に同期化させる処理を行うことを特徴とする。

20

30

40

【0007】

第3の本発明は、音声・映像送信端末が並行的にネットワークに送信した音声信号及び映像信号が与えられ、上記音声・映像送信端末からの音声信号及び映像信号を同期化させる音声・映像受信端末において、(1)対向する上記音声・映像送信端末が、(1-1)自己が有するローカル時計の時刻を含む音声についての音声ローカル時刻通知信号を上記音声・映像受信端末に向けて送信する音声ローカル時刻通知信号送信手段と、(1-2)ネットワークで調停済みの自己が有する調停済時計の時刻を含む音声についての調停済時

50

刻通知信号を上記音声・映像受信端末に向けて送信する音声調停済時刻通知信号送信手段と、(1-3)自己が有するローカル時計の時刻を含む映像についてのローカル時刻通知信号を上記音声・映像受信端末に向けて送信する映像ローカル時刻通知信号送信手段と、(1-4)ネットワークで調停済みの自己が有する調停済時計の時刻を含む映像についての調停済時刻通知信号を上記音声・映像受信端末に向けて送信する映像調停済時刻通知信号送信手段とを備えたものであり、(2)自端末は、(2-1)上記音声調停済時刻通知信号が与えられたときに、ネットワークで調停済みの自己が有する調停済時計での受信時刻と、その通知信号での時刻と、上記音声ローカル時刻通知信号の当該音声・映像受信端末での再生タイミングに基づいて得た音声についての処理待ち時間とから、映像についてのオフセット時間を得る映像オフセット時間取得手段と、(2-2)上記映像調停済時刻通知信号が与えられたときに、自己が有する上記調停済時計での受信時刻を上記オフセット時間分だけ修正すると共に、上記映像ローカル時刻通知信号が与えられたときに、自己が有するローカル時計の時刻を、その通知信号での時刻に修正する映像時刻修正手段と、(2-3)上記映像時刻修正手段が修正した2種類の時刻に基づいて、上記音声・映像送信端末が送信した映像信号を上記音声・映像送信端末が送信した音声信号に同期化させる処理を行う映像処理手段とを備えることを特徴とする。

10

【発明の効果】

【0008】

本発明によれば、受信側における音声及び映像の同期確立を簡易に実現できる音声・映像同期方法、音声・映像同期システム及び音声・映像受信端末を提供できる。

20

【発明を実施するための最良の形態】

【0009】

(A)主たる実施形態

以下、本発明による音声・映像同期方法及び音声・映像同期システムを、テレビ電話システムに適用した一実施形態を、図面を参照しながら説明する。

【0010】

(A-1)実施形態の構成

この実施形態のテレビ電話システム1は、図2に示すように、2台のテレビ電話端末2A及び2Bが、IP網3を介して、音声信号及び映像信号を授受するものである。音声信号及び映像信号は、別個の packets によって授受され、何らの対策がなされないと仮定した場合には、音声及び映像の同期がとれないシステムを前提としている。

30

【0011】

この実施形態の場合、音声信号についても映像信号についても、端末2A及び2B間にて、音声及び映像用のUDP(User Datagram Protocol)セッションをオープンし、それぞれに対応するRTP(Real-Time Transport Protocol)/RTCP(RTP Control Protocol)パケットを送受信するようになされている。なお、RTPパケットやRTCPパケットのパケット送出間隔は、各端末2A、2Bによって異なるものである。

【0012】

各テレビ電話端末2A、2Bは、専用端末として構築されたものであっても良く、また、IPテレビ電話用のソフトウェアをパソコン(PC)などに実装することで構築されたもの(IPソフトフォン)であってても良く、機能的には、図3に示す詳細構成を有する。

40

【0013】

図3において、テレビ電話端末2(2A、2B)は、音声を捕捉するマイクロフォン10、マイクロフォン10からの音声信号を圧縮する音声圧縮部11、音声圧縮部11からの音声ストリームを送信処理するストリーミング送信部12を有すると共に、被写体を撮像するカメラ13、カメラ13からの映像信号を圧縮する映像圧縮部14、映像圧縮部14からの映像ストリームを送信処理するストリーミング送信部15を有する。例えば、音声圧縮部11、ストリーミング送信部12、映像圧縮部14及びストリーミング送信部15は、IPソフトフォンの場合には、ソフトウェアを中心として構成される。

50

【 0 0 1 4 】

また、テレビ電話端末 2 は、音声ストリームを受信処理するストリーミング受信部 1 6、ストリーミング受信部 1 6 から出力された圧縮音声信号を解凍する音声解凍部 1 7、音声解凍部 1 7 から出力された音声信号を発音出力するスピーカ部 1 8 を有すると共に、映像ストリームを受信処理するストリーミング受信部 1 9、ストリーミング受信部 1 9 から出力された圧縮映像信号を解凍する映像解凍部 2 0、映像解凍部 2 0 から出力された映像信号を表示出力するディスプレイ部 2 1 を有する。例えば、ストリーミング受信部 1 6、音声解凍部 1 7、ストリーミング受信部 1 9 及び映像解凍部 2 0 は、IP ソフトフォンの場合には、ソフトウェアを中心として構成される。

【 0 0 1 5 】

なお、音声及び映像の圧縮方式は任意であるが、例えば、MPEG 4 (Moving Picture Experts Group 4) を適用し得る。

【 0 0 1 6 】

さらに、テレビ電話端末 2 は、対向するテレビ電話端末とのセッションの確立や切断などの処理を行うシグナリング処理部 2 2、音声や映像などのメディアやシグナリングから IP パケットを組み立てて送信する IP 送信部 2 3、到来した IP パケットを受信処理してメディアやシグナリングを取り出す IP 受信部 2 4 を有する。なお、シグナリング処理部 2 2 は、例えば、セッションの確立や切断などの処理を行う SIP 処理部や、ヘッダ情報などの解析するための SDP 処理部などを有する。また、周知のように、IP パケットの中には UDP パケットを収容したものがあり、UDP パケットには RTP パケットや RTPCP パケットが適宜収容されている。

【 0 0 1 7 】

IP 送信部 2 3 及び IP 受信部 2 4 はそれぞれ、音声用の IP パケットと映像用の IP パケットとを別個に取り扱うものである。

【 0 0 1 8 】

この実施形態の場合、IP 送信部 2 3 及び IP 受信部 2 4 に関連して、RTP/RTCP 通信部 2 5 が設けられている。RTP/RTCP 通信部 2 5 は、図 1 に示すような機能的構成を有する。

【 0 0 1 9 】

RTP/RTCP 通信部 2 5 は、ローカルタイマ部 3 0、NTP (Network Time Protocol) タイマ部 3 1、RTP/RTCP 送信部 3 2、RTP/RTCP 受信・同期処理部 3 3 を有する。

【 0 0 2 0 】

ローカルタイマ部 3 0 は、他の装置などと連携しないで時刻を計時している、当該テレビ電話端末 2 における独自のタイマである。NTP タイマ部 3 1 は、NTP に従って、時刻を計時しているタイマである。

【 0 0 2 1 】

RTP/RTCP 送信部 3 2 は、当該テレビ電話端末 2 の音声や映像についてそれぞれ定まっているパケット送出間隔で、音声や映像の RTP パケットや RTCP パケットを IP 送信部 2 3 に与えて送信させるものであり、RTP ヘッダには、ローカルタイマ部 3 0 の計時時刻による RTP タイムスタンプを挿入し、RTCP ヘッダには、ローカルタイマ部 3 0 の計時時刻による RTP タイムスタンプ及び NTP タイマ部 3 1 の計時時刻による NTP タイムスタンプを挿入するものである (このような挿入動作は、RFC に準拠した動作である)。

【 0 0 2 2 】

RTP/RTCP 受信・同期処理部 3 3 は、S1 処理部 3 3 - 1 ~ S5 処理部 3 3 - 5 を備え、対向するテレビ電話端末から到来した RTP パケットや RTCP パケットをスタックしながら、それら RTP パケットや RTCP パケットに基づいて (RTP パケットや RTCP パケットを解析して)、映像及び音声を同期化させるものである。例えば、音声解凍部 1 7 及び映像解凍部 2 0 から出力される音声信号及び映像信号を同期化させるもの

10

20

30

40

50

である。

【 0 0 2 3 】

映像及び音声の同期化は、後述する動作説明で明らかにするが、この実施形態の場合、音声を基準に、映像を同期化させている。映像信号についてはフレーム間符号化が適応的になされ、その圧縮後のデータ量の変化が大きい（言い換えるとジッタが大きい）が、音声信号の場合には、圧縮後のデータ量が固定の圧縮方式が適用されていることが多く、ジッタの変化が少ないため、音声を基準とすることとした。

【 0 0 2 4 】

(A - 2) 実施形態の動作

次に、実施形態のテレビ電話システム 1 の動作、特に、受信側における音声と映像とを同期化させる動作（実施形態の音声・映像同期方法）を、図 4 を参照しながら詳述する。図 4 は、S 1 処理部 3 3 - 1 ~ S 5 処理部 3 3 - 5 の処理をそれぞれ示す説明図である。テレビ電話システム 1 では、2 つのテレビ電話端末 2 A 及び 2 B 間で双方向通信が実行されるが、以下では、テレビ電話端末 2 A が送信側、テレビ電話端末 2 B が受信側として説明する。

10

【 0 0 2 5 】

送信側のテレビ電話端末 2 A は、上述したように、当該テレビ電話端末 2 A について定まっている、音声及び映像のそれぞれのパケット送出間隔で、音声や映像の RTP パケットや RTCP パケットを送信させる。この送信の際、RTP ヘッダには RTP タイムスタンプを挿入し、RTCP ヘッダには RTP タイムスタンプ及び NTP タイムスタンプを挿入する。

20

【 0 0 2 6 】

(処理 S 1)

テレビ電話端末 2 B の S 1 処理部 3 3 - 1 は、音声に係る RTCP パケットを受信した場合には、定義された共通関数などを使用し、当該テレビ電話端末 2 B が受信した時刻 LOCAL_NTP_A を求める。この受信時刻 LOCAL_NTP_A は、NTP を適用したものであるため、概ね正確なものである。

【 0 0 2 7 】

(処理 S 2)

テレビ電話端末 2 B の S 2 処理部 3 3 - 2 は、処理 S 1 を起動させた RTCP パケット受信時における音声処理部（ストリーミング受信部 1 6 及び音声解凍部 1 7 ）の内部の処理待ちバッファを確認し、当該 RTCP パケットヘッダ内の RTP タイムスタンプに対応する RTP パケットが、どのタイミングで再生されるのかを計算する。すなわち、RTP タイムスタンプに対応する RTP パケットの処理待ち時間 Queue_A を計算する。

30

【 0 0 2 8 】

(処理 S 3)

テレビ電話端末 2 B の S 3 処理部 3 3 - 3 は、処理 S 1 及び S 2 で得られた 2 つのパラメータ LOCAL_NTP_A、Queue_A と、受信した RTCP パケット内の NTP タイムスタンプ REMOTE_NTP_A とから、(1) 式に示す時刻差分 OFFSET_A を算出する。

40

【 0 0 2 9 】

OFFSET_A =

$$LOCAL_NTP_A - REMOTE_NTP_A + Queue_A \quad (1)$$

受信した RTCP パケット内の NTP タイムスタンプ REMOTE_NTP_A は、送信側のテレビ電話端末 2 A が送信した時刻を表しているため、(1) 式における LOCAL_NTP_A - REMOTE_NTP_A の部分は、IP 網 3 での伝搬遅延に相当する。この伝搬遅延に、受信側のテレビ電話端末 2 B における処理待ち時間 Queue_A を加算しているため、(1) 式に示す時刻差分 OFFSET_A は、送信側のテレビ電話端末 2 A が RTP パケットを送信してから、受信側のテレビ電話端末 2 B がその RTP パケットの処理を開始するまでの時刻差を表している。

50

【 0 0 3 0 】

以上から明らかなように、時刻差分 $OFFSET_A$ は、RTP パケットを受信する毎に新しいものに更新される。

【 0 0 3 1 】

(処理 S 4)

テレビ電話端末 2 B の S 4 処理部 3 3 - 4 は、映像に係る RTP パケットを受信した場合には、受信した RTP ヘッダ内の対向するテレビ電話端末 2 A の送出時刻を表す NTP タイムスタンプ $REMOTE_NTP_V$ と RTP タイムスタンプ $REMOTE_RTP_V$ とに対し、(2) 式及び (3) 式を実行し、その実行後の値を、映像処理部 (ストリーミング受信部 1 9 及び映像解凍部 2 0) に渡す。

10

【 0 0 3 2 】

$$LOCAL_NTP_V = REMOTE_NTP_V + OFFSET_A \quad (2)$$

$$LOCAL_RTP_V = REMOTE_RTP_V \quad (3)$$

(2) 式に示す時刻 $LOCAL_NTP_V$ は、音声に係る RTP パケットが送信されてから受信側でその RTP パケットの処理が開始されるまでの時間差分 $OFFSET_A$ を、送信側のテレビ電話端末 2 A が映像に係る RTP パケットを送信した時点に加算したものであり、時刻 $LOCAL_NTP_V$ に、受信した映像に係る RTP パケットの処理を開始すると仮定すると、送信側のテレビ電話端末 2 A が RTP パケットを送信した時点から、受信側のテレビ電話端末 2 B が RTP パケットの処理を開始するまでの時間が、音声と映像とで同一時間となる。

20

【 0 0 3 3 】

(3) 式に示す時刻 $LOCAL_RTP_V$ は、受信側のテレビ電話端末 2 B における独自タイマ (ローカルタイマ部) での時刻 (受信時刻) を表しており、(3) 式は、送信側のテレビ電話端末 2 A における独自タイマ (ローカルタイマ部) での時刻 (送信時刻) $REMOTE_RTP_V$ に、受信側のテレビ電話端末 2 B における独自タイマでの時刻 (受信時刻) をすることを意味しており、双方の独自タイマが連携されたものとなる。

【 0 0 3 4 】

(処理 S 5)

テレビ電話端末 2 B の S 5 処理部 3 3 - 5 は、(2) 式及び (3) 式で求められた情報に基づき、映像処理部が同期処理を行うようにさせる。映像処理部は、対応する RTP パケットと、当該テレビ電話端末 2 B の時刻情報を元に、映像の再生タイミングを同期させ、これにより、音声を基準とした映像同期を実現する。

30

【 0 0 3 5 】

(A - 3) 実施形態の効果

上記実施形態によれば、受信側における音声及び映像の同期のために、送信側が行う処理は、RFC 準拠の RTP / RTP パケットを送信するだけで良く、既存の処理と同様である。そのため、送信側及び受信側の構成を備える端末の場合、音声及び映像の同期確立のための構成を小規模のものとすることができ、また、音声及び映像の同期確立のための処理を簡易にし得る。

【 0 0 3 6 】

また、上記実施形態によれば、受信側にて、音声、映像の解凍 (デコード) 処理がリンクして動作しなくても、音声を基準とした映像同期を実現することができる。

40

【 0 0 3 7 】

(B) 他の実施形態

上記実施形態では、本発明をテレビ電話システム (ピアツーピア接続の一例) に適用したものを示したが、テレビ会議システムなどの 1 対多通信にも、本発明を適用でき、音声を基準とした映像同期を実現することができる。

【 0 0 3 8 】

また、上記実施形態では、映像についての RTP パケットや RTP パケットの処理タイミングを調整することにより、音声を基準とした映像同期を実現するものを示したが、

50

【 図 4 】



フロントページの続き

審査官 坂本 聡生

- (56)参考文献 特開2003-304511(JP,A)
特開平10-136308(JP,A)
特開平10-271482(JP,A)
特表2010-512688(JP,A)
特開2006-310964(JP,A)
特開平07-050818(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04N	7/14	-	7/173
H04N	7/00	-	7/088
H04L	12/00	-	12/26
H04L	12/50	-	12/66
H04M	3/00		
H04M	3/16	-	3/20
H04M	3/38	-	3/58
H04M	7/00	-	7/16
H04M	11/00	-	11/10
H04W	40/34		