(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6852478号
(P6852478)

(45)発行日　令和3年3月31日(2021.3.31)　(24)登録日　令和3年3月15日(2021.3.15)

(51)Int.Cl.		F I		
G10L 15/22	(2006.01)	G10L	15/22	460Z
G10L 15/30	(2013.01)	G10L	15/30	
G10L 21/055	(2013.01)	G10L	21/055	
G10L 15/00	(2013.01)	G10L	15/00	200A
H04N 7/15	(2006.01)	H04N	7/15	150

請求項の数 7　(全 17 頁)　最終頁に続く

(21)出願番号　特願2017-48205(P2017-48205)
(22)出願日　平成29年3月14日(2017.3.14)
(65)公開番号　特開2018-151533(P2018-151533A)
(43)公開日　平成30年9月27日(2018.9.27)
審査請求日　令和2年1月16日(2020.1.16)

(73)特許権者　000006747
株式会社リコー
東京都大田区中馬込１丁目３番６号
(74)代理人　100107766
弁理士　伊東　忠重
(74)代理人　100070150
弁理士　伊東　忠彦
(72)発明者　中島　章敬
東京都大田区中馬込１丁目３番６号　株式会社リコー内

審査官　菊池　智紀

最終頁に続く

(54)【発明の名称】通信端末、通信プログラム及び通信方法

(57)【特許請求の範囲】
【請求項１】
集音装置により集音した音声データの送受信を行う通信端末であって、
前記音声データを前記通信端末とネットワークを介して接続された音声認識装置へ送信し、前記音声認識装置から前記音声データの音声認識結果であるテキストデータを受信する通信部と、
前記音声データを再生し、前記音声データが再生されている期間中に前記テキストデータを表示装置に表示させる出力部と、
前記音声データに対する加工を行うか否かを判定する加工判定部と、
前記音声データに対し、前記音声データの再生時間を延ばす加工を行う音声加工部と、を有し、
前記加工判定部は、前記音声データの受信が開始されてから、前記テキストデータを受信するまでの期間が、所定の期間よりも長いとき、前記音声加工部による加工を行うものと判定する、通信端末。
【請求項２】
前記通信部により、前記音声データの受信が開始されてから、前記テキストデータの受信が完了するまでの期間、受信した前記音声データを保持させるバッファ処理部を有し、
前記出力部は、
前記テキストデータの受信が完了した後に、前記音声データの再生と、前記テキストデータの表示と、を同時に開始する、請求項１記載の通信端末。

【請求項３】
前記音声加工部は、
加工後の前記音声データの再生時間が、前記音声データの受信が開始されてから、前記テキストデータを受信するまでの期間よりも長くなるように、前記音声データの加工を行い、
前記出力部は、
加工後の前記音声データが再生されている期間中に前記テキストデータを表示装置に表示させる、請求項１又は２記載の通信端末。
【請求項４】
前記所定の期間は、
前記音声データの受信が開始されてから、前記テキストデータを受信するまでの期間の平均である、請求項２又は３記載の通信端末。
【請求項５】
前記音声データは、
前記表示装置に表示された画像の画像データ、前記表示装置に対して、入力された文字や画像を示すストローク情報の少なくとも何れかを含むコンテンツデータに含まれる、請求項１乃至４の何れか一項に記載の通信端末。
【請求項６】
集音装置により集音した音声データの送受信を行う通信端末による通信方法であって、前記通信端末が、
前記音声データを前記通信端末とネットワークを介して接続された音声認識装置へ送信し、前記音声認識装置から前記音声データの音声認識結果であるテキストデータを受信する手順と、
前記音声データを再生し、前記音声データが再生されている期間中に前記テキストデータを表示装置に表示させる手順と、
前記音声データに対する加工を行うか否かを判定する手順と、
前記音声データに対し、前記音声データの再生時間を延ばす加工を行う手順と、を有し、
前記判定する手順において、前記音声データの受信が開始されてから、前記テキストデータを受信するまでの期間が、所定の期間よりも長いとき、前記加工を行う手順による加工を行うものと判定する、通信方法。
【請求項７】
集音装置により集音した音声データの送受信を行う通信端末において実行される通信プログラムであって、
前記音声データを前記通信端末とネットワークを介して接続された音声認識装置へ送信し、前記音声認識装置から前記音声データの音声認識結果であるテキストデータを受信する処理と、
前記音声データを再生し、前記音声データが再生されている期間中に前記テキストデータを表示装置に表示させる処理と、
前記音声データに対する加工を行うか否かを判定する処理と、
前記音声データに対し、前記音声データの再生時間を延ばす加工を行う処理と、通信端末に実行させ、
前記判定する処理において、前記音声データの受信が開始されてから、前記テキストデータを受信するまでの期間が、所定の期間よりも長いとき、前記加工を行う処理による加工を行うものと判定する、処理を前記通信端末に実行させる、通信プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、通信端末、通信プログラム及び通信方法に関する。
【背景技術】

【０００２】
従来から、音声データをテキストデータに変換する音声認識サービスを利用して、音声と対応したテキストデータを表示させる技術が普及している。
【０００３】
具体的には、例えば、会議システムにおいて、音声認識サービスにより発言者の音声データをテキストデータに変換し、聞き手が使用しているディスプレイに、発言者の音声データに対応したテキストデータを表示させる技術が知られている（特許文献１）。
【発明の概要】
【発明が解決しようとする課題】
【０００４】
従来の会議システムでは、音声データが入力された後に、この音声データが音声認識サービスに送信されてテキストデータに変換されるため、音声データが再生されるタイミングとテキストデータが表示されるタイミングにずれが生じる。
【０００５】
そのため、従来では、再生される音声データの内容と、表示されるテキストデータの内容とが対応せず、聞き手に違和感を与える、音声データの内容の理解を妨げる等の可能性があった。
【０００６】
開示の技術は、上記事情に鑑みてなされたものであり、再生中の音声データと、表示されるテキストデータとを対応させることを目的としている。
【課題を解決するための手段】
【０００７】
開示の技術は、集音装置により集音した音声データの送受信を行う通信端末であって、前記音声データを前記通信端末とネットワークを介して接続された音声認識装置へ送信し、前記音声認識装置から前記音声データの音声認識結果であるテキストデータを受信する通信部と、前記音声データを再生し、前記音声データが再生されている期間中に前記テキストデータを表示装置に表示させる出力部と、前記音声データに対する加工を行うか否かを判定する加工判定部と、前記音声データに対し、前記音声データの再生時間を延ばす加工を行う音声加工部と、を有し、前記加工判定部は、前記音声データの受信が開始されてから、前記テキストデータを受信するまでの期間が、所定の期間よりも長いとき、前記音声加工部による加工を行うものと判定する。
【発明の効果】
【０００８】
再生中の音声データと、表示されるテキストデータとを対応させることができる。
【図面の簡単な説明】
【０００９】
【図１】通信システムのシステム構成の一例を示す図である。
【図２】第一の実施形態の通信システムの動作の概略を説明するシーケンス図である。
【図３】比較例となる通信システムの動作の概略を説明するシーケンス図である。
【図４】第一の実施形態の通信端末のハードウェア構成の一例を示す図である。
【図５】第一の実施形態の通信端末の有する各装置の機能を説明する図である。
【図６】第一の実施形態の通信システムの動作を説明するシーケンス図である。
【図７】第一の実施形態の通信端末の動作を説明する図である。
【図８】第一の実施形態の通信端末の表示装置の表示例を示す図である。
【図９】第二の実施形態の通信端末の有する各装置の機能を説明する図である。
【図１０】第二の実施形態の通信システムの動作を説明するシーケンス図である。
【図１１】第二の実施形態の通信端末の動作を説明するフローチャートである。
【図１２】第二の実施形態の通信端末の動作を説明する図である。
【発明を実施するための形態】
【００１０】

（第一の実施形態）
以下に図面を参照して、第一の実施形態について説明する。図１は、通信システムのシステム構成の一例を示す図である。
【００１１】
本実施形態の通信システム１００は、通信端末２００－１、２００－２と、サーバ装置３００と、を有する。通信システム１００において、通信端末２００－１、２００－２、サーバ装置３００のそれぞれは、ネットワークＮを介して接続される。また、本実施形態の通信システム１００は、ネットワークＮを介して、音声データをテキストデータに変換する音声認識装置４００と接続される。
【００１２】
本実施形態の通信システム１００において、通信端末２００－１、２００－２は、例えば電子黒板等であっても良く、サーバ装置３００は、例えばテレビ会議を実現するためのテレビ会議用のサーバ装置等であっても良い。以下の説明において、通信端末２００－１、２００－２のそれぞれを区別しない場合には、通信端末２００と呼ぶ。
【００１３】
本実施形態の通信端末２００は、マイク等の集音装置を有しており、集音装置によって集音された音声データを、サーバ装置３００と、音声認識装置４００とに送信する。また、本実施形態の通信端末２００は、手書き入力された文字や画像等を示すストローク情報、画面をキャプチャした画像データ等を、サーバ装置３００に送信する。さらに、本実施形態の通信端末２００は、カメラ等の撮像装置を有しており、撮像装置によって撮像された画像データを、サーバ装置３００に送信する。
【００１４】
本実施形態の音声認識装置４００は、例えば、人工知能により提供されるサービスである。音声認識装置４００は、受信した音声データを音声認識機能によりテキストデータとし、サーバ装置３００へ送信する。尚、本実施形態の音声データは、通信端末２００の周辺で発話した人の声や、人の声以外の様々な音等、集音装置が集音した全ての音の音声データである。したがって、本実施形態では、通信端末２００の周辺で発話した人の声を示す発話データは、音声データの一部である。
【００１５】
本実施形態のサーバ装置３００は、受信したストローク情報や画像データ、音声データ等を格納する。また、本実施形態のサーバ装置３００は、音声認識装置４００から送信されたテキストデータを、音声データと対応付けて格納する。以下の説明では、通信端末２００からサーバ装置３００に送信される各種のデータをコンテンツデータと呼ぶ。したがって、本実施形態のコンテンツデータは、音声データ、画像データ、ストローク情報等を含む。
【００１６】
また、本実施形態のサーバ装置３００は、例えば、ある会議において通信端末２００が使用された場合、会議名と、会議中に取得したコンテンツデータと、音声データから変換されたテキストデータとが対応付けられて格納されても良い。言い換えれば、サーバ装置３００では、通信端末２００から取得したコンテンツデータが、会議毎に格納されても良い。
【００１７】
本実施形態の通信システム１００では、例えば、通信端末２００－１の利用者と、通信端末２００－２の利用者とにより、テレビ会議を行うことができる。この場合、サーバ装置３００は、通信端末２００－１、２００－２のそれぞれから入力された情報を、通信端末２００－１、２００－２のそれぞれの画面に表示させ、情報を共有させる。
【００１８】
具体的には、通信システム１００の通信端末２００は、一方の通信端末２００において撮像された画像データと集音された音声データとを、サーバ装置３００を介して、他方の通信端末２００に送信する。

【００１９】
他方の通信端末２００は、サーバ装置３００から受信した画像データを表示装置に表示させ、音声データを再生する。また、他方の通信端末２００は、受信した音声データを音声認識装置４００へ送信してテキストデータとし、このテキストデータを表示装置に表示させる。
【００２０】
ここで、本実施形態の通信端末２００では、サーバ装置３００から受信した音声データが再生されている期間中に、この音声データの音声認識結果であるテキストデータを表示装置に表示させる。
【００２１】
本実施形態では、この処理により、通信端末２００において、再生される音声データと、表示されるテキストデータとを対応させる。
【００２２】
尚、図１では、通信端末２００の一例として、電子黒板としているが、これに限定されない。本実施形態の通信端末２００は、集音装置と表示装置を有しており、外部の装置（サーバ装置３００、音声認識装置４００）と通信を行うことができる端末であれば良い。具体的には、本実施形態の通信端末２００は、例えば、一般的なコンピュータ、タブレット型端末、スマートフォン等がある。また、その他にも、各種の電子機器に本実施形態を適用することができる。
【００２３】
以下に、図２及び図３を参照して、本実施形態の通信システム１００の動作の概略について説明する。図２は、第一の実施形態の通信システムの動作の概略を説明するシーケンス図である。
【００２４】
本実施形態の通信システム１００において、通信端末２００－１は、撮像された画像データと集音された音声データとをサーバ装置３００へ送信する（ステップＳ２０１）。サーバ装置３００は、受信した画像データと音声データとを、通信端末２００－２に送信する（ステップＳ２０２）。
【００２５】
通信端末２００－２は、画像データと音声データとを受信すると、音声データを音声認識装置４００へ送信する（ステップＳ２０３）。また、通信端末２００－２は、受信した画像データと音声データとを、一時的に保持する（ステップＳ２０４）。
【００２６】
続いて、通信端末２００－２は、音声認識装置４００から、ステップＳ２０３で送信した音声データの音声認識結果であるテキストデータを受信すると（ステップＳ２０５）、画像データと音声データを再生させ、受信したテキストデータを表示させる（ステップＳ２０６）。
【００２７】
このように、本実施形態の通信端末２００では、音声データを受信した場合には、この音声データの音声認識結果のテキストデータを取得するまで、音声データの再生を行わずに待機する。
【００２８】
以下に、図３を参照して、本実施形態が適用されない通信システムの動作を説明する。図３は、比較例となる通信システムの動作の概略を説明するシーケンス図である。
【００２９】
図３のステップＳ３０１からステップＳ３０３までの処理は、図２のステップＳ２０１からステップＳ２０３までの処理と同様であるから、説明を省略する。
【００３０】
通信端末２－２は、ステップＳ３０３において音声データを音声認識装置４００に送信すると、サーバ装置３００から受信した画像データと音声データとを再生する（ステップ

Ｓ３０４）。続いて、通信端末２－２は、音声認識装置４００からテキストデータを受信し（ステップＳ３０５）、受信したテキストデータを表示装置に表示させる（ステップＳ３０６）。

【００３１】

このように、図３の例では、通信端末２－２は、音声データを受信すると、テキストデータに変換される前に音声データの再生を開始する。したがって、図３の例では、テキストデータを受信して表示するまでの間に、音声データの再生が終了する可能性がある。この場合、通信端末２－２では、表示装置に表示されたテキストデータは、すでに再生が終了した音声データと対応するものとなる。よって、図３の例では、再生される音声データと表示されるテキストデータとが対応しない。

【００３２】

これに対し、図２に示す本実施形態の通信端末２００－２は、音声データを受信した後に、この音声データの音声認識結果であるテキストデータを受信するまで、音声データの再生を保留する。そして、通信端末２００－２は、テキストデータを受信した後に、音声データの再生とテキストデータの表示とを行う。したがって、本実施形態によれば、音声データが再生されている期間中に、この音声データの音声認識結果であるテキストデータを表示させることができ、音声データとテキストデータとを対応させることができる。

【００３３】

以下に、本実施形態の通信端末２００について説明する。図４は、第一の実施形態の通信端末のハードウェア構成の一例を示す図である。

【００３４】

本実施形態の通信端末２００は、入力装置２１と、表示装置２２と、外部Ｉ／Ｆ２３と、通信Ｉ／Ｆ２４と、ＲＯＭ２５（Read Only Memory）とを有する。また、本実施形態の通信端末２００は、ＲＡＭ（Random Access Memory）２６と、ＣＰＵ（Central Processing Unit）２７と、ＨＤＤ（Hard Disk Drive）２８と、集音装置２９と、撮像装置３０と、を有する。これらの各ハードウェアは、それぞれがバスＢ１で接続されている。

【００３５】

入力装置２１は、タッチパネル等であり、ユーザによる各種操作（例えば、音声テキスト変換（日本語）や音声テキスト変換（英語）等の機能の選択操作）を入力するのに用いられる。表示装置２２は、ディスプレイ等であり、各種情報（例えば、音声テキスト変換（日本語）による変換結果を示すテキストや音声データ等）を表示する。尚、本実施形態では、タッチパネルが、入力装置と２１と表示装置２２の両方を兼ねていても良い。

【００３６】

外部Ｉ／Ｆ２３は、外部装置とのインターフェースである。外部装置には、記録媒体２３ａ等がある。これにより、通信端末２００は、外部Ｉ／Ｆ２３を介して、記録媒体２３ａ等の読み取りや書き込みを行うことができる。なお、記録媒体２３ａには、例えば、ＵＳＢメモリやＣＤ、ＤＶＤ、ＳＤメモリカード等がある。

【００３７】

通信Ｉ／Ｆ２４は、通信端末２００をネットワークＮ１等に接続するためのインターフェースである。これにより、通信端末２００は、通信Ｉ／Ｆ２４を介して、他の装置（と通信を行うことができる。

【００３８】

ＨＤＤ２８は、プログラムやデータを格納している不揮発性の記憶装置である。ＨＤＤ２８に格納されるプログラムやデータには、通信端末２００全体を制御する基本ソフトウェアであるＯＳ（Operating System）、ＯＳ上において各種機能を提供するアプリケーションプログラム等がある。

【００３９】

また、ＨＤＤ２８は、格納しているプログラムやデータを所定のファイルシステム及び／又はＤＢ（データベース）により管理している。なお、通信端末２００は、ＨＤＤ２８に代えて、記録媒体としてフラッシュメモリを用いるドライブ装置（例えばソリッドステ

ートドライブ：ＳＳＤ）を有していても良い。
【００４０】
ＲＯＭ２５は、電源を切ってもプログラムやデータを保持することができる不揮発性の半導体メモリである。ＲＯＭ２５には、通信端末２００の起動時に実行されるＢＩＯＳ（Basic Input/Output System）、ＯＳ設定、及びネットワーク設定等のプログラムやデータが格納されている。ＲＡＭ２６は、プログラムやデータを一時保持する揮発性の半導体メモリである。
【００４１】
ＣＰＵ２７は、ＲＯＭ２５やＨＤＤ２８等の記憶装置からプログラムやデータをＲＡＭ２６上に読み出し、処理を実行することで、通信端末２００全体の制御や機能を実現する演算装置である。
【００４２】
集音装置２９は、例えばマイクロフォン（マイク）等であり、通信端末２００の周囲の音を集音する。
【００４３】
撮像装置３０は、例えばカメラ等であり、通信端末２００の周辺の画像を撮像する。具体的には、例えば、撮像装置３０は、通信端末２００を用いて会議等を行っている様子等を撮像する。
【００４４】
本実施形態の通信端末２００は、図２に示すハードウェア構成を有することにより、後述するような各種処理を実現できる。
【００４５】
次に、図５を参照して、本実施形態の通信端末２００の機能について説明する。図５は、第一の実施形態の通信端末の有する各装置の機能を説明する図である。
【００４６】
本実施形態の通信端末２００の機能は、ＣＰＵ２７がＲＡＭ２６等に格納されたプログラムを読み出して実行することで実現される。
【００４７】
本実施形態の通信端末２００は、集音部２１０、入力部２２０、出力部２３０、コンテンツ保持部２４０、バッファ処理部２５０、通信部２６０を有する。
【００４８】
集音部２１０は、集音装置２９に入力された音声を音声データとして取得する。入力部２２０は、通信端末２００の有するタッチパネル（入力装置２１、表示装置２２）に対して手書き入力された文字や画像を示すストローク情報や、タッチパネルに表示された画像の画像データ等を取得する。尚、本実施形態のストローク情報とは、タッチパネルに対して手書き入力が行われた場合の、利用者による一画毎の軌跡を示す点群の座標報である。
【００４９】
入力部２２０は、撮像装置３０により撮影された画像データを取得する。尚、本実施形態の画像データは、動画データと静止画データの両方を含む。
【００５０】
出力部２３０は、音声データや画像データを出力する。具体的には、出力部２３０は、例えば、表示装置２２に対して画像データを表示させたり、音声データを再生させたりする。
【００５１】
コンテンツ保持部２４０は、バッファ処理部２５０の指示により、音声データを一時的に保持する。本実施形態のコンテンツ保持部２４０は、例えば、通信部２６０によりサーバ装置３００から受信したコンテンツデータを保持しても良いし、集音部２１０と入力部２２０により取得されたコンテンツデータを保持しても良い。また、コンテンツ保持部２４０は、コンテンツデータに含まれる音声データのみを保持しても良い。
【００５２】

バッファ処理部２５０は、音声データが音声認識装置４００に送信された場合、音声認識装置４００から音声認識結果のテキストデータを受信するまで、コンテンツデータをコンテンツ保持部２４０に保持させる。

【００５３】

通信部２６０は、サーバ装置３００及び音声認識装置４００との通信を行う。具体的には、通信部２６０は、集音部２１０、入力部２２０により取得したコンテンツデータをサーバ装置３００へ送信する。また、通信部２６０は、集音部２１０により取得した音声データを音声認識装置４００へ送信し、音声認識結果のテキストデータを受信する。また、通信部２６０は、他の通信端末２００から送信されたコンテンツデータを、サーバ装置３００を介して受信する。

【００５４】

次に、図６を参照して、本実施形態の通信システム１００の動作について説明する。図６は、第一の実施形態の通信システムの動作を説明するシーケンス図である。

【００５５】

図６では、通信端末２００－１が取得したコンテンツデータを、サーバ装置３００を介して通信端末２００－２が受信する場合の動作を示している。

【００５６】

本実施形態の通信システム１００において、通信端末２００－１は、集音部２１０－１により音声データを取得すると、通信部２６０－１へ渡す（ステップＳ６０１）。また、通信端末２００－１の入力部２２０－２により画像データを取得すると、通信部２６０－１へ渡す（ステップＳ６０２）。通信端末２００－１の通信部２６０－１は、音声データと画像データとを含むコンテンツデータをサーバ装置３００へ送信する（ステップＳ６０３）。

【００５７】

サーバ装置３００は、このコンテンツデータを、通信端末２００－２へ送信する（ステップＳ６０４）。

【００５８】

通信端末２００－２は、コンテンツデータを受信すると、通信部２６０－２により、コンテンツデータに含まれる音声データを音声認識装置４００へ送信する（ステップＳ６０５）。また、通信部２６０－２は、バッファ処理部２５０に、コンテンツデータを渡す（ステップＳ６０６）。

【００５９】

バッファ処理部２５０－２は、コンテンツデータを受けて、コンテンツデータをコンテンツ保持部２４０－２へ保持させる（ステップＳ６０７）。

【００６０】

続いて、通信端末２００－２は、通信部２６０－２により、音声認識装置４００から、ステップＳ６０５で送信した音声データの音声認識結果のテキストデータを受信する（ステップＳ６０８）。

【００６１】

続いて、通信部２６０－２は、バッファ処理部２５０－２に受信したテキストデータを渡す（ステップＳ６０９）。バッファ処理部２５０－２は、テキストデータを受けると、コンテンツ保持部２４０－２からコンテンツデータを取得する（ステップＳ６１０）。

【００６２】

そして、バッファ処理部２５０－２は、出力部２３０－２に、コンテンツデータとテキストデータとを渡す（ステップＳ６１１）。出力部２３０－２は、コンテンツデータとテキストデータとを同時に出力させる（ステップＳ６１２）。

【００６３】

以下に、図７を参照して、本実施形態の通信端末２００の動作について、さらに具体的に説明する。図７は、第一の実施形態の通信端末の動作を説明する図である。

【００６４】

図７の例では、本実施形態の通信端末２００－２では、タイミングＴ１において、サーバ装置３００からコンテンツデータの受信を開始し、タイミングＴ２において、コンテンツデータの受信を完了したものとする。

【００６５】

このとき、通信端末２００－２は、タイミングＴ１において、コンテンツデータの受信を開始した直後からコンテンツデータの再生を行わず、バッファ処理部２５０－２により、コンテンツ保持部２４０－２にコンテンツデータを保持させる。

【００６６】

そして、通信端末２００－２は、タイミングＴ２において、コンテンツデータの受信が完了し、タイミングＴ３において、テキストデータの受信が完了すると、コンテンツデータの再生と、テキストデータの表示を同時に開始する。したがって、本実施形態では、タイミングＴ１からタイミングＴ３までの期間Ｋ１３は、コンテンツ保持部２４０－２にコンテンツデータが保持されるコンテンツ保持期間となる。

【００６７】

したがって、通信端末２００－２では、タイミングＴ１からタイミングＴ３の間に、コンテンツデータに含まれる音声データを音声認識装置４００へ送信し、音声認識結果のテキストデータを受信する。尚、本実施形態では、タイミングＴ２において、コンテンツデータの受信が完了した後に、音声認識装置４００への音声データの送信を開始しても良いし、タイミングＴ１においてコンテンツデータの受信を開始したときから、音声認識装置４００への音声データの送信を開始しても良い。

【００６８】

図７の例では、コンテンツデータの再生時間は、タイミングＴ３からタイミングＴ５までの期間Ｋ３５であり、テキストデータの表示時間は、タイミングＴ３からタイミングＴ４までの期間Ｋ３４となる。

【００６９】

したがって、本実施形態の通信端末２００では、音声データを含むコンテンツデータが再生されている期間中に、この音声データと対応するテキストデータが表示装置２２に表示させることができる。

【００７０】

このように、本実施形態によれば、再生中の音声データと、表示されるテキストデータとを対応させることができるため、聞き手に対して違和感を与えることがない。また、本実施形態によれば、音声データの再生中にテキストデータを表示することによって、音声データの内容の理解を支援することができる。

【００７１】

図８は、第一の実施形態の通信端末の表示装置の表示例を示す図である。図８に示す画面８１は、例えば、拠点Ａに設置された通信端末２００－１と、拠点Ｂに設置された通信端末２００－２とを用いてテレビ会議を行った場合の、通信端末２００－１、２００－２のそれぞれの表示装置２２に表示される画面の例である。

【００７２】

画面８１は、通信端末２００－１の入力部２２０－１により取得された画像データが表示される表示領域８２、通信端末２００－２の入力部２２０－２により取得された画像データが表示される表示領域８３を含む。また、画面８１には、通信端末２００－１の集音部２１０－１と、通信端末２００－２の集音部２１０－２のそれぞれから取得された音声データから変換されたテキストデータが表示される表示領域８４を含む。

【００７３】

本実施形態では、例えば、表示領域８３において表示された利用者の画像データと音声データが再生されている期間中に、表示領域８４において、この音声データと対応するテキストデータが表示される。したがって、本実施形態によれば、音声データの内容を示すテキストデータが、音声データの再生中に表示されることになる。このため、本実施形態によれば、画面８１を閲覧している利用者に対して、音声データの再生のタイミングと、

テキストデータの表示のタイミングのずれを感じさせることがなく、操作性を向上させることができる。

【００７４】

尚、本実施形態では、主に通信端末２００－２の動作を説明したが、通信端末２００－１と通信端末２００－２は、同様の構成を有するものであり、通信端末２００－１も、通信端末２００－２と同様の動作を行うものである。

【００７５】

（第二の実施形態）

以下に図面を参照して、第二の実施形態について説明する。第二の実施形態では、コンテンツデータを受信してから音声データを再生するまでのコンテンツ保持期間に応じて、音声データに対する加工を行う点が、第一の実施形態と相違する。よって、以下の第二の実施形態の説明では、第一の実施形態との相違点についてのみ説明し、第一の実施形態と同様の機能構成を有するものには、第一の実施形態の説明で用いた符号を付与し、その説明を省略する。

【００７６】

図９は、第二の実施形態の通信端末の有する各装置の機能を説明する図である。本実施形態の通信端末２００Ａは、第一の実施形態の通信端末２００の有する各部に加え、保持期間取得部２７０、加工判定部２８０、音声加工部２９０を有する。

【００７７】

本実施形態の保持期間取得部２７０は、通信部２６０がコンテンツデータの受信を開始してから、コンテンツデータに含まれる音声データの音声認識結果のテキストデータを受信するまでのコンテンツ保持期間を算出して取得し、記憶する。

【００７８】

加工判定部２８０は、コンテンツ保持期間に基づき、音声データに対する加工を行うか否かを判定する。具体的には、加工判定部２８０は、コンテンツ保持期間が所定の期間より長い期間であるか否かを判定し、所定の期間より長い期間である場合には、音声データに対する加工を行うものと判定する。また、加工判定部２８０は、コンテンツ保持期間が所定の期間以内である場合には、音声データに対する加工はせずに、コンテンツ保持部２４０により、音声データを含むコンテンツデータを保持するものと判定する。

【００７９】

また、本実施形態の所定の期間とは、例えば、過去にコンテンツデータを受信した際のコンテンツ保持期間の平均である。本実施形態では、例えば、加工判定部２８０が、保持期間取得部２７０によりコンテンツ保持期間が記憶される度に、所定の期間となるコンテンツ保持期間の平均を算出し、保持していても良い。

【００８０】

音声加工部２９０は、前回のコンテンツデータの受信の際に、保持期間取得部２７０により記憶されたコンテンツ保持期間が、予め設定された所定の期間より長い期間であった場合に、コンテンツデータに含まれる音声データを引き伸ばす加工を行う。具体的には、音声加工部２９０は、音声データである波形を編集することで、音声データの再生時間を延ばすようにしても良い。

【００８１】

尚、本実施形態の音声加工部２９０は、コンテンツデータに含まれる画像データが静止画の画像データである場合には、音声データについてのみ、加工を行っても良い。また、本実施形態の音声加工部２９０は、コンテンツデータに含まれる画像データが動画データであるは場合には、動画データに対しても、音声データに対する加工と同様の加工を行う。

【００８２】

例えば、コンテンツデータが、動画データであり、フレームレートが３０［ｆｐｓ］である場合には、このフレームレートを１５［ｆｐｓ］とすれば、コンテンツデータの再生時間を２倍にすることができる。

【００８３】
以下に、図１０を参照して、本実施形態の通信システム１００の動作について説明する。図１０は、第二の実施形態の通信システムの動作を説明するシーケンス図である。
【００８４】
図１０では、通信端末２００Ａ－１からサーバ装置３００に送信されたコンテンツデータを通信端末２００Ａ－２が受信した際に、コンテンツ保持期間が所定の期間より長い期間であった場合の動作を示している。
【００８５】
図１０のステップＳ１００１とステップＳ１００２の処理は、図２のステップＳ２０１とステップＳ２０２の処理と同様であるから、説明を省略する。
【００８６】
通信端末２００Ａ－２は、通信部２６０－２により、コンテンツデータを受信すると、保持期間取得部２７０により、前回の保持されているコンテンツ保持期間を取得する（ステップＳ１００３）。
【００８７】
続いて、通信端末２００Ａ－２は、加工判定部２８０により、取得したコンテンツ保持期間が所定の期間より長い期間であると判定された場合に、音声加工部２９０により、コンテンツデータに含まれる音声データを加工する（ステップＳ１００４）。このとき、音声加工部２９０は、加工後の音声データを再生したときの再生時間が、コンテンツ保持期間よりも長くなるように、音声データを引き延ばす加工を行う。
【００８８】
続いて、通信端末２００Ａ－２は、音声データを音声認識装置４００に送信する（ステップＳ１００５）。また、通信端末２００Ａ－２は、加工された音声データと画像データとの再生を開始する（ステップＳ１００６）。
【００８９】
続いて、通信端末２００Ａ－２は、音声認識装置４００より、音声データの音声認識結果のテキストデータを受信し（ステップＳ１００７）、このテキストデータを表示装置２２に表示させる（ステップＳ１００８）。
【００９０】
以上のように、本実施形態では、コンテンツ保持期間の長さに応じて、音声データの再生時間が、コンテンツ保持期間よりも長くなるように加工する。したがって、本実施形態によれば、加工後の音声データの再生中に、この音声データと対応するテキストデータを表示させることができる。
【００９１】
以下に、図１１を参照して、本実施形態の通信端末２００Ａの全体の動作を説明する。図１１は、第二の実施形態の通信端末の動作を説明するフローチャートである。本実施形態の通信端末２００Ａは、コンテンツデータを受信する度に、図１１の処理を実行する。
【００９２】
本実施形態の通信端末２００Ａは、通信部２６０により、コンテンツデータの受信を開始すると（ステップＳ１１０１）、保持期間取得部２７０により、前回のコンテンツデータの受信において記憶されたコンテンツ保持期間を取得する（ステップＳ１１０２）。
【００９３】
続いて、通信端末２００Ａは、加工判定部２８０により、コンテンツ保持期間が所定の期間より長い期間であるか否かを判定する（ステップＳ１１０３）。ステップＳ１１０３において、コンテンツ保持期間が所定の期間よりも長い場合、通信端末２００Ａは、後述するステップＳ１１０９へ進む。
【００９４】
ステップＳ１１０３において、コンテンツ保持期間が所定の期間以内である場合、通信端末２００Ａは、バッファ処理部２５０により、コンテンツデータをコンテンツ保持部２４０に保持させる（ステップＳ１１０４）。続いて、通信端末２００Ａは、通信部２６０

により、音声データを音声認識装置４００へ送信し（ステップＳ１１０５）、音声認識装置４００からテキストデータを受信する（ステップＳ１１０６）。

【００９５】

続いて通信端末２００Ａは、出力部２３０により、コンテンツデータを再生し、テキストデータを表示させる（ステップＳ１１０７）。続いて、通信端末２００Ａは、保持期間取得部２７０により、ステップＳ１１０１でコンテンツデータの受信を開始してから、ステップＳ１１０６でテキストデータを受信するまでのコンテンツ保持期間を取得して記憶し（ステップＳ１１０８）、処理を終了する。

【００９６】

ステップＳ１１０３において、コンテンツ保持期間が所定の期間より長い場合、通信端末２００Ａは、音声加工部２９０により、音声データの再生時間がコンテンツ保持期間よりも長くなるように、音声データを引き伸ばす加工を行う（ステップＳ１１０９）。

【００９７】

続いて、通信端末２００Ａは、通信部２６０より、加工していない音声データを音声認識装置４００へ送信する（ステップＳ１１１０）。また、通信端末２００Ａは、出力部２３０により、加工された音声データと、コンテンツデータに含まれる画像データと、の再生を開始する（ステップＳ１１１１）。

【００９８】

続いて、通信端末２００Ａは、通信部２６０により、音声認識装置４００からテキストデータを受信し（ステップＳ１１１２）、出力部２３０により、受信したテキストデータを表示装置２２に表示させ（ステップＳ１１１３）、ステップＳ１１０８へ進む。

【００９９】

以下に、図１２を参照して、実施形態の通信端末２００Ａの動作について、さらに説明する。図１２は、第二の実施形態の通信端末の動作を説明する図である。

【０１００】

図１２の例では、本実施形態の通信端末２００Ａ－２では、タイミングＴ１において、サーバ装置３００からコンテンツデータの受信を開始し、タイミングＴ２において、コンテンツデータの受信を完了したものとする。また、図１２では、通信端末２００Ａ－２の保持期間取得部２７０により記憶されたコンテンツ保持期間が所定の期間より長いときの、通信端末２００Ａ－２の動作を示している。

【０１０１】

図１２の例では、タイミングＴ１において、コンテンツデータの受信を開始した直後からコンテンツデータの保持を行わず、コンテンツデータの再生を開始する。このとき、コンテンツデータに含まれる音声データは、コンテンツ保持期間よりも再生時間が長くなるように、加工されている。

【０１０２】

図１２において、通信端末２００Ａ－２は、タイミングＴ２において、コンテンツデータの受信が完了する。また、通信端末２００Ａ－２は、タイミングＴ３において、テキストデータを受信すると、タイミングＴ３からタイミングＴ４までの期間Ｋ３４において、テキストデータを表示させる。

【０１０３】

このとき、コンテンツデータは、まだ再生中である。図１２の例では、通信端末２００Ａ－２は、タイミングＴ５において、加工後の音声データを含むコンテンツデータの再生が完了する。したがって、加工後の音声データを含むコンテンツデータの再生時間は、タイミングＴ１からタイミングＴ５までの期間Ｋ１５となる。

【０１０４】

したがって、本実施形態によれば、テキストデータが表示される期間Ｋ３４は、加工後の音声データを含むコンテンツデータの再生時間である期間Ｋ１５に含まれることになる。

【０１０５】

このように、本実施形態では、音声データ（コンテンツデータ）を引き延ばすことで、音声データの再生中に、この音声データの音声認識結果であるテキストデータが表示されるようにすることができる。

【０１０６】

以上、各実施形態に基づき本発明の説明を行ってきたが、上記実施形態に示した要件に本発明が限定されるものではない。これらの点に関しては、本発明の主旨をそこなわない範囲で変更することができ、その応用形態に応じて適切に定めることができる。

【符号の説明】

【０１０７】

１００　通信システム
２００、２００Ａ　通信端末
２１０　集音部
２２０　入力部
２３０　出力部
２４０　コンテンツ保持部
２５０　バッファ処理部
２６０　通信部
２７０　保持期間取得部
２８０　加工判定部
２９０　音声加工部
３００　サーバ装置
４００　音声認識装置

【先行技術文献】

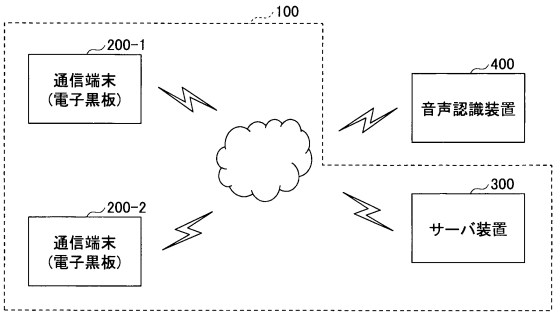
【特許文献】

【０１０８】

【特許文献１】特開２０１１－１８２１２５号公報

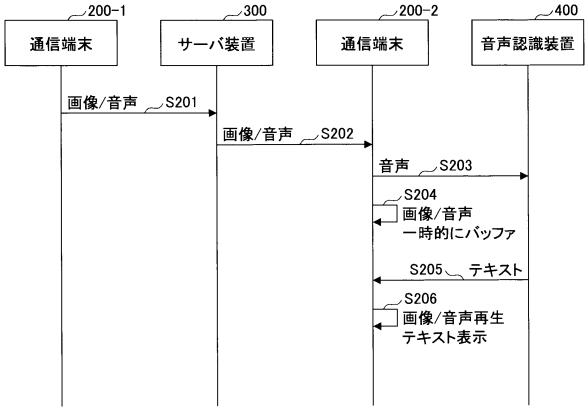
【図１】

通信システムのシステム構成の一例を示す図



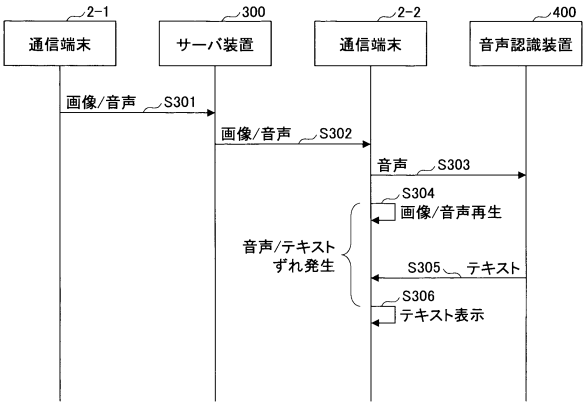
【図２】

第一の実施形態の通信システムの動作を説明するシーケンス図



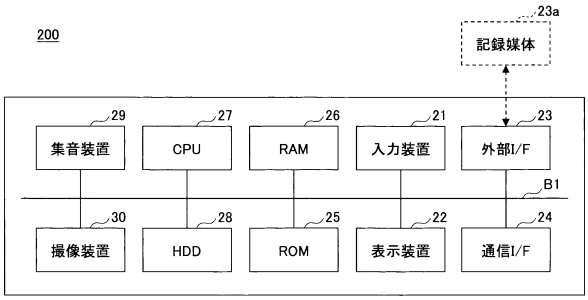
【図３】

比較例となる通信システムの動作を説明するシーケンス図



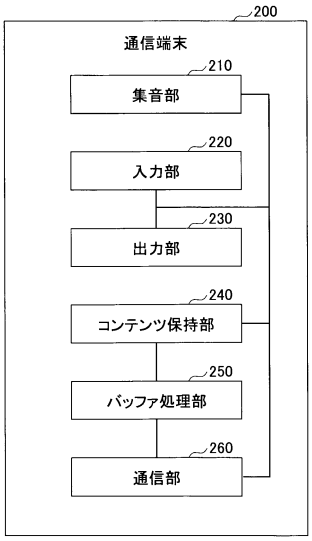
【図４】

第一の実施形態の通信端末のハードウェア構成の一例を示す図



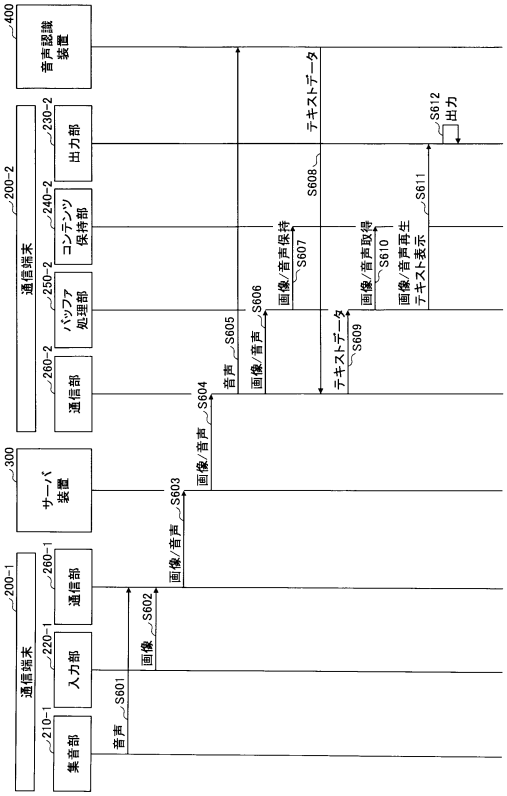
【図５】

第一の実施形態の通信端末の有する各装置の機能を説明する図



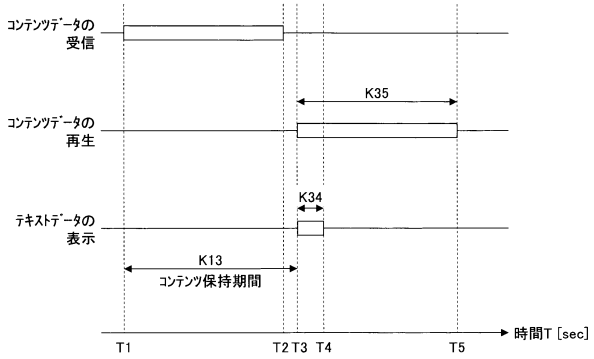
【図６】

第一の実施形態の通信システムの動作を説明するシーケンス図



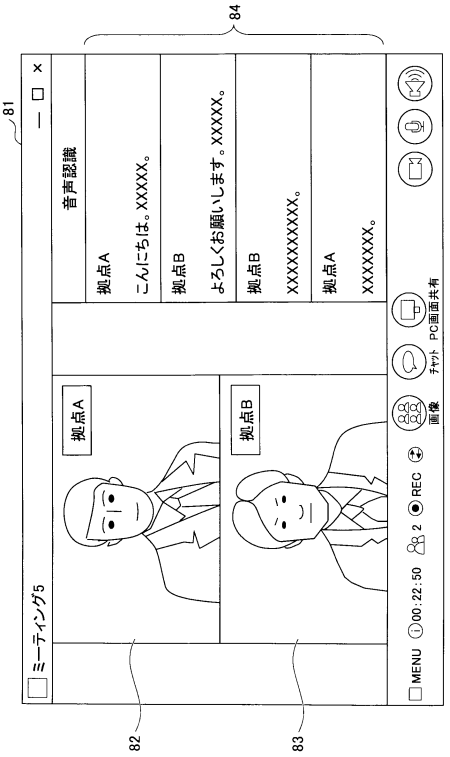
【図７】

第一の実施形態の通信端末の動作を説明する図



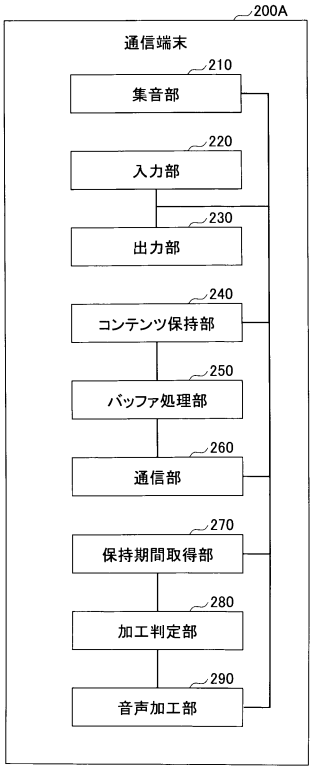
【図８】

第一の実施形態の通信端末の表示装置の表示例を示す図



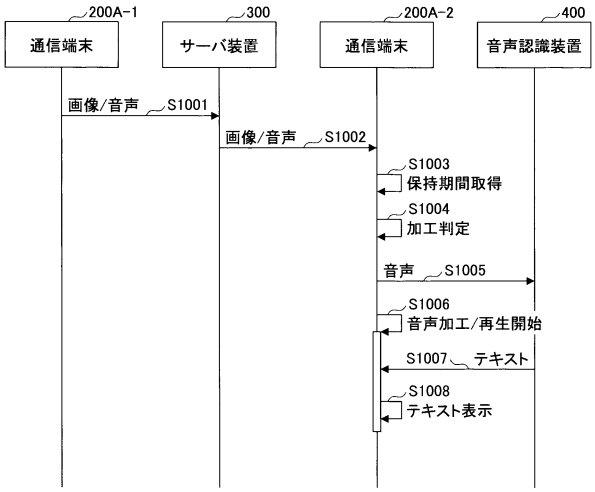
【図９】

第二の実施形態の通信端末の有する各装置の機能を説明する図



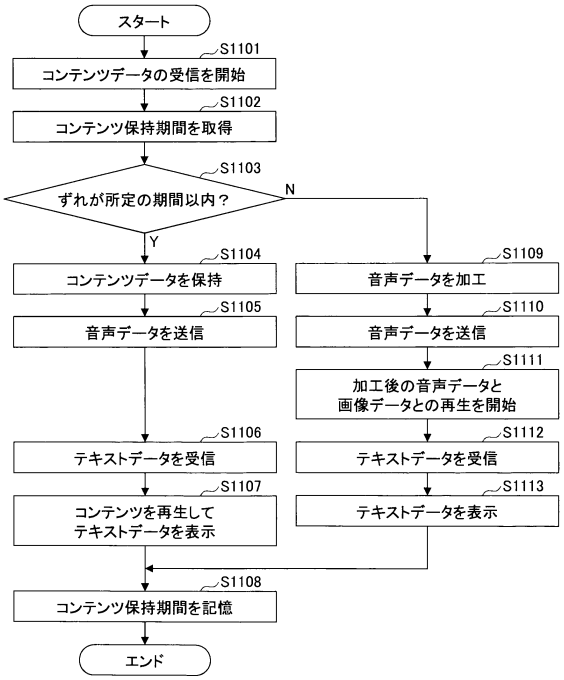
【図１０】

第二の実施形態の通信システムの動作を説明するシーケンス図



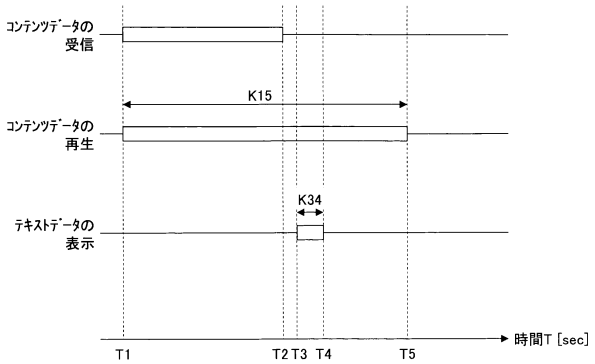
【図１１】

第二の実施形態の通信端末の動作を説明するフローチャート



【図１２】

第二の実施形態の通信端末の動作を説明する図



フロントページの続き

(51)Int.Cl.			FI		
H04M	*3/42*	*(2006.01)*	H04M	3/42	R

(56)参考文献 特開２０１０－２３０９４８（ＪＰ，Ａ）
特開２００５－６４６００（ＪＰ，Ａ）
特開２００３－２０９６００（ＪＰ，Ａ）
特開２００４－３０４６０１（ＪＰ，Ａ）
特開２０１２－１２９９５０（ＪＰ，Ａ）
特開２００８－６６８６６（ＪＰ，Ａ）
米国特許出願公開第２０１４／０１９２１３８（ＵＳ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１５／００－１７／２６，
２１／００－２１／０５７
Ｈ０４Ｍ　１／００－ ３／６４
Ｈ０４Ｎ　７／１４－ ７／１５