

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(10) 国際公開番号

WO 2012/014275 A1

PCT

(43) 国際公開日

2012年2月2日(02.02.2012)

- (51) 国際特許分類:
H04L 12/56 (2006.01) H04M 11/06 (2006.01)
- (21) 国際出願番号: PCT/JP2010/062558
- (22) 国際出願日: 2010年7月26日(26.07.2010)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人(米国を除く全ての指定国について): パイオニア株式会社(PIONEER CORPORATION) [JP/JP]; 〒2120031 神奈川県川崎市幸区新小倉1番1号 Kanagawa (JP).
- (72) 発明者; および
- (75) 発明者/出願人(米国についてのみ): 亀谷 隆真(KAMETANI, Ryushin) [JP/JP]; 〒2120031 神奈川県川崎市幸区新小倉1-1 パイオニア株式会社 川崎事業所内 Kanagawa (JP).
- (74) 代理人: 江上 達夫, 外(EGAMI, Tatsuo et al.); 〒1040031 東京都中央区京橋一丁目16番10号

オークビル京橋3階 東京セントラル特許事務所内 Tokyo (JP).

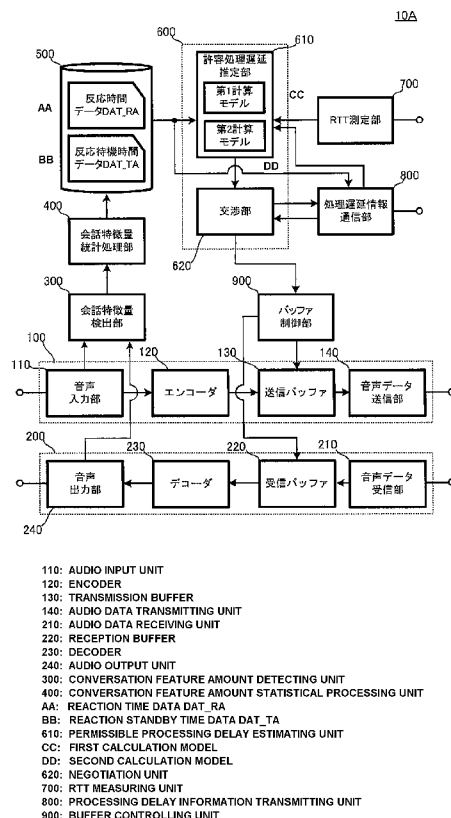
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF,

[続葉有]

(54) Title: AUDIO TRANSMITTING/RECEIVING DEVICE, AUDIO TRANSMITTING/RECEIVING SYSTEM AND SERVER DEVICE

(54) 発明の名称: 音声送受信装置、音声送受信システム及びサーバ装置

[図2]



(57) Abstract: This audio transmitting/receiving device, audio transmitting/receiving system and server device provides optimal conversation quality suited to the situation of a user engaged in a conversation. The audio transmitting/receiving device (10A) comprises: transmitting means (100, 200) for transmitting and receiving conversation data over a network (20) to and from an audio transmitting/receiving device (10B) provided at a base other than the home base, among multiple bases, said conversation data representing the content of a conversation and including at least audio data; a conversation feature amount acquiring means (300) for acquiring conversation feature amounts for a home base user (A) located at the home base and another base user (B) located at the other base, respectively; a processing delay amount determining means (600) for determining a total amount of processing delay amounts in the entire audio transmitting/receiving system on the basis of conversation feature amounts acquired; and a processing delay amount controlling means (600) for controlling the processing delay amounts so as to satisfy the total amount determined.

(57) 要約: 会話に参加するユーザの事情に即した最適な会話品質を提供する。音声送受信装置(10A)は、複数の拠点のうち自拠点を除く他拠点に設置される音声送受信装置(10B)との間で、ネットワーク(20)を介して、少なくとも音声データを含む、会話の内容を表す会話データの送受信を行う通信手段(100、200)と、自拠点に存在する自拠点ユーザ(A)及び他拠点に存在する他拠点ユーザ(B)の各々について会話特徴量を取得する会話特徴量取得手段(300)と、取得された会話特徴量に基づいて、会話データの処理遅延量の、音声送受信システム全体における総量を決定する処理遅延量決定手段(600)と、決定された総量が満たされるように前記処理遅延量を制御する処理遅延量制御手段(600)とを具備する。

WO 2012/014275 A1



BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, 添付公開書類:
SN, TD, TG).

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称：

音声送受信装置、音声送受信システム及びサーバ装置

技術分野

- [0001] 本発明は、例えば遠隔地のユーザ相互間において会話を円滑に進行させることが可能な音声送受信装置、音声送受信システム及びサーバ装置の技術分野に関する。

背景技術

- [0002] この種の目的を有するものとして、例えば特許文献1には、リアルタイム音声再生装置が提案されている。この装置によれば、音声の欠落を会話として成立する最大の音声欠落率以内に収め、音声遅延を会話として成立する最大の遅延時間以内に収めることにより、全体の会話品質を確保することが可能であるとされている。
- [0003] また、特許文献2には、遅延ゆらぎを吸収するためのバッファを、有音区間でのバッファアンダーフロー発生回数等に基づいて無音区間に制御するための、バッファ監視部とバッファ制御部とを備えた構成が開示されている。

先行技術文献

特許文献

- [0004] 特許文献1：特開2002-222324号公報
特許文献2：特開平11-215812号公報

発明の概要

発明が解決しようとする課題

- [0005] 一般的には、遅延時間は、大きい程、音切れの発生や誤りの発生を抑制し得るから、会話の音声品質を向上させ得る。一方で、遅延時間が大きければ、一方のユーザが、自身の発話終了時点以降、相手の発話開始を認識するまでの時間が長くなるため、会話の円滑な進行が妨げられ易い。従って、特許文献1の装置のように、遅延時間を会話として成立する最大の遅延時間以内

に収める旨の技術思想は、音声品質と円滑な進行との妥協点を見出し得る点において有益である。

[0006] ところで、会話が成立する最大の遅延時間は、例えば、会話に参加するユーザ各々の性格又は身体的若しくは精神的負荷状態、或いはその時点のユーザの個別具体的な事情等に応じて、如何様にも変化し得る。例えば、最大の遅延時間は、性格的にせっかちなユーザであれば短く、のんびりしたユーザであれば長く、イライラしているユーザであれば短く、リラックスしたユーザであれば長く、体調の悪いユーザであれば体調に応じて長くも短くもなる。また、それとは別に、何らかの事情で急いでいれば当然ながら短くなる。

[0007] 特許文献 1 に開示された装置を適用する場合、会話として成立する最大の遅延時間を事前に設定する必要があるが、このように如何様にも変化し得る性質を有する最大の遅延時間を、予め実験的に、経験的に或いは理論的に確定させておくことは、実践上困難を極める。従って、特許文献 1 で述べられるところの最大の遅延時間は、必ずしも真に会話を成立させるための最大の遅延時間とはならない。このため、特許文献 1 の装置を会話に適用したとしても、ユーザによっては、快適性の低下を招く可能性がある。或いは、逆に遅延時間が短過ぎて、不要な音声品質の低下を招く可能性がある。即ち、特許文献 1 の装置には、ユーザ側の事情に適応する術を有さぬことに起因して、遅延時間が最適化され難いという技術的問題点がある。

[0008] また、このような問題点は、特許文献 2 の装置のように、遅延揺らぎの吸収による音声パケットの廃棄率低減を目的としたバッファ制御をなし得たところで、何ら変わらず生じ得る。

[0009] 本発明は、係る問題点に鑑みてなされたものであり、会話に参加するユーザの事情に即した最適な会話品質を提供可能な音声送受信装置、音声送受信システム及びサーバ装置を提供することを課題とする。

課題を解決するための手段

[0010] 上述した課題を解決するため、請求の範囲第 1 項の音声送受信装置は、各々がネットワークに收容される、複数の拠点に設置された音声送受信装置を

含み、該音声送受信装置を介して、前記複数の拠点に夫々存在するユーザ相互間で会話を成立させることが可能な音声送受信システムにおける、前記音声送受信装置であって、前記複数の拠点のうち自拠点を除く他拠点に設置される前記音声送受信装置との間で、前記ネットワークを介して、少なくとも音声データを含む、前記会話の内容を表す会話データの送受信を行う通信手段と、前記自拠点到存在する自拠点ユーザ及び前記他拠点到存在する他拠点ユーザの各々について、前記会話における発話のタイミングに関連する所定の会話特徴量を取得する会話特徴量取得手段と、前記取得された会話特徴量に基づいて、その大小が前記会話の時間遅延量の大小に夫々対応し且つ前記会話の音声品質の高低に夫々対応する前記会話データの処理遅延量の、前記音声送受信システム全体における総量を決定する処理遅延量決定手段と、前記決定された総量が満たされるように前記処理遅延量を制御する処理遅延量制御手段とを具備することを特徴とする。

[0011] 上述した課題を解決するため、請求の範囲第14項の音声送受信システムは、各々がネットワークに收容される、複数の拠点到設置された音声送受信装置を含み、該音声送受信装置を介して、前記複数の拠点到夫々存在するユーザ相互間で会話を成立させることが可能な音声送受信システムであって、前記音声送受信装置は、前記複数の拠点のうち自拠点を除く他拠点到設置される前記音声送受信装置との間で、前記ネットワークを介して、少なくとも音声データを含む、前記会話の内容を表す会話データの送受信を行う通信手段と、前記自拠点到存在する自拠点ユーザ及び前記他拠点到存在する他拠点ユーザの各々について、前記会話における発話のタイミングに関連する所定の会話特徴量を取得する会話特徴量取得手段と、前記取得された会話特徴量に基づいて、その大小が前記会話の時間遅延量の大小に夫々対応し且つ前記会話の音声品質の高低に夫々対応する前記会話データの処理遅延量の、前記音声送受信システム全体における総量を決定する処理遅延量決定手段と、前記決定された総量が満たされるように前記処理遅延量を制御する処理遅延量制御手段とを具備することを特徴とする。

[0012] 上述した課題を解決するため、請求の範囲第15項のサーバ装置は、各々がネットワークに收容される、サーバ装置と、複数の拠点に設置された、該複数の拠点のうち自拠点を除く他拠点に設置される音声送受信装置との間で、前記ネットワークを介して、少なくとも音声データを含む、前記会話の内容を表す会話データの送受信を行う通信手段を具備する音声送受信装置とを含み、該音声送受信装置を介して、前記複数の拠点に夫々存在するユーザ相互間で会話を成立させることが可能な音声送受信システムにおける、前記サーバ装置であって、前記複数の音声送受信装置から、前記ネットワークを介して、前記自拠点到存在する自拠点ユーザ及び前記他拠点到存在する他拠点ユーザの各々についての前記会話における発話のタイミングに関連する所定の会話特徴量を取得する会話特徴量取得手段と、前記取得された会話特徴量に基づいて、その大小が前記会話の時間遅延量の大小に夫々対応し且つ前記会話の音声品質の高低に夫々対応する前記会話データの処理遅延量の、前記音声送受信システム全体における総量を決定する処理遅延量決定手段と、前記複数の音声送受信装置に対し前記ネットワークを介して前記決定された総量を告知する告知手段とを具備することを特徴とする。

図面の簡単な説明

[0013] [図1]本発明の第1実施例に係る遠隔会議システムの構成を概念的に表してなる概略構成図である。

[図2]図1の遠隔会議システムにおける音声送受信装置の構成を概念的に表してなるブロック図である。

[図3]図2の音声送受信装置において実行されるバッファ容量制御のフローチャートである。

[図4]反応時間の概念を説明するタイミングチャートである。

[図5]図2の音声送受信装置において実行される反応時間算出処理のフローチャートである。

[図6]図2の音声送受信装置において実行される反応待機時間算出処理のフローチャートである。

[図7] 第1計算モデルの概念を説明するタイミングチャートである。

[図8] 第2計算モデルの概念を説明するタイミングチャートである。

[図9] 本発明の第2実施例に係る遠隔会議システムの構成を概念的に表してなる概略構成図である。

発明を実施するための形態

[0014] <音声送受信装置の実施形態>

[0015] 本発明の音声送受信装置に係る実施形態は、各々がネットワークに收容される、複数の拠点に設置された音声送受信装置を含み、該音声送受信装置を介して、前記複数の拠点に夫々存在するユーザ相互間で会話を成立させることが可能な音声送受信システムにおける、前記音声送受信装置であって、前記複数の拠点のうち自拠点を除く他拠点に設置される前記音声送受信装置との間で、前記ネットワークを介して、少なくとも音声データを含む、前記会話の内容を表す会話データの送受信を行う通信手段と、前記自拠点に存在する自拠点ユーザ及び前記他拠点に存在する他拠点ユーザの各々について、前記会話における発話のタイミングに関連する所定の会話特徴量を取得する会話特徴量取得手段と、前記取得された会話特徴量に基づいて、その大小が前記会話の時間遅延量の大小に夫々対応し且つ前記会話の音声品質の高低に夫々対応する前記会話データの処理遅延量の、前記音声送受信システム全体における総量を決定する処理遅延量決定手段と、前記決定された総量が満たされるように前記処理遅延量を制御する処理遅延量制御手段とを具備する。

[0016] 実施形態に係る音声送受信装置は、相異なる複数の拠点間で各拠点に存在するユーザ同士の会話を成立させ得るシステムとしての音声送受信システムを構築する、一の音声送受信装置である。実施形態に係る音声送受信装置は、ネットワークに常時、或いは何らかの条件（尚、この種の条件は、如何様にも限定されない）が満たされた場合に限定的に收容される構成となっている。

[0017] 実施形態に係る「ネットワーク」とは、例えばWAN（Wide Area Network）網、LAN（Local Area Network）網、又はこれらWAN網又はLAN網

を介して或いは電話回線、A D S L (Asymmetric Digital Subscriber Line) 又は光ファイバーケーブル等を介して適宜に接続されるインターネット網等の各種データ通信網を包括する概念である。

[0018] 実施形態に係る音声送受信装置は、通信手段を有しており、この通信手段の作用により、自拠点とは異なる他拠点（即ち、会話の相手側の拠点である）に設置される音声送受信装置との間で、ネットワークを介した会話データの送受信を行うことが可能である。尚、他拠点に設置される音声送受信装置との間の、当該会話データを始めとする各種情報、データ、又はデータファイル等の送受信形態は、例えば然るべきサーバ装置を介したものであってもよいし、例えばP 2 P (Peer To Peer) 等サーバ装置を介さないものであってもよい。必然的に、音声送受信システムの構成要素もまた多義的であってよい。

[0019] 尚、自拠点とは、実施形態に係る音声送受信装置が設置される拠点を意味するものであって、ある特定の拠点を限定的に意味するものではない。

[0020] 実施形態における「会話データ」とは、夫々相異なる拠点に存在するユーザ相互間の会話の成立に必要となる或いは意義を与え得るデータを包括する概念であり、特に、少なくとも音声データを含むものとして規定される。会話データには、例えば、この他に画像データや映像データ等が含まれてもよい。尚、「会話」とは、音声を伴った、ユーザ相互間の意思伝達行為を意味し、発生するシチュエーションは多岐にわたる。例えば、実施形態における「会話」とは、日常会話の他に、打合せや会議等における参加者各々の発話等が好適に含まれる。

[0021] この会話データのうち、他拠点に設置された音声送受信装置へ送信されるデータは、例えば、マイク等の集音手段を介して集音された自拠点ユーザのアナログ音声データを、エンコーダ等を介して符号化してなるデータであってもよい。このデータは、他拠点側に設置された音声送受信装置において受信され（即ち、他拠点側では受信データとなる）、例えばデコーダ等を介して復号化され、最終的にスピーカ等の出力手段を介してアナログ音声として

他拠点ユーザの受話に供され得る。発話側と受話側とが入れ替わっても同様である。実施形態に係る音声送受信装置では、例えばこのように会話データの送受信が繰り返されることにより、相異なる拠点に存在するユーザ相互間で会話を成立させることができる。

[0022] 実施形態に係る音声送受信装置は、会話データの処理遅延量の、音声送受信システム全体における総量（以下、適宜「システム総処理遅延量」と称する）を決定する処理遅延量決定手段と、この決定されたシステム総処理遅延量が満たされるように当該処理遅延量を制御する処理遅延量制御手段とを具備する。

[0023] ここで、「会話データの処理遅延量」とは、複数の拠点の各々において音声送受信装置が会話データを処理するにあたって生じる遅延の量であって、特に、その大小が、自拠点ユーザと他拠点ユーザとの会話の時間遅延の量の大小に夫々対応し、且つその大小が、自拠点ユーザと他拠点ユーザとの会話の音声品質の高低に夫々対応する、可制御性を有する遅延の量を意味する。システム総処理遅延量とは、発話側と受話側とにおける当該処理遅延量の総和である。従って、例えば音切れや符号化誤りの少ない、高い音声品質を得ようとする場合、システム総処理遅延量は大きい方がよいことになる。

[0024] ところで、会話の時間遅延は、音声品質とは異なる次元で会話の品質に影響する。より具体的には、会話の時間遅延量が大きくなり過ぎると、ユーザ相互間で同一の時間軸を共有することが難しくなり、会話のリアルタイム性が低下する。このようなリアルタイム性の低下は、それ自体が快適性を低下させユーザに不快感を惹起する。また、このようなリアルタイム性の低下は、発話してから初めて相手が発話中であることを確認する等、円滑性の低下による二次的な会話品質の低下を招き、一種の悪循環を招来し易い。

[0025] このように、音声送受信システムにおいては、会話の音声品質と快適性とはトレードオフの関係となる。従って、これら双方に影響するシステム総処理遅延量を最適化することが必要となる。

[0026] 尚、実施形態に係る会話データには、ユーザの表情を特定するための画像

データや映像データが含まれ得るが、この種の画像データや映像データの送受信は、それ自体が高負荷処理であり、音声と画像又は映像を同期出力しようとするれば、会話の時間遅延量は増大する。一方で、これらを独立に制御すれば、表情と音声との同期が崩れるため、快適性の低下を抑制する旨の効果は限定的となる。即ち、上述の問題点は、画像又は映像を併用することにより解消される性質のものではない。

[0027] そこで、実施形態に係る音声送受信装置では、会話特徴量取得手段により、自拠点に存在する自拠点ユーザ及び他拠点に存在する他拠点ユーザの各々について所定の会話特徴量が取得される。処理遅延量決定手段は、この取得された会話特徴量に基づいてシステム総処理遅延量を決定する構成となっている。

[0028] 会話特徴量とは、会話における発話のタイミングに関連する物理量、制御量或いは規格化された各種の指標値等を意味する。会話における発話のタイミングは、好適な一形態として、相手の発話終了後に発話を開始するタイミングと、自分の発話終了後に相手が無反応であると判断して再度発話を開始するタイミングとに大別される。これらは、いずれも会話の基本的な進行リズムを規定する重要な要素であり、最適なシステム総処理遅延量と有意な関係を有し得る。例えば、発話のタイミングが早い（遅い）ユーザに対する最適なシステム総処理遅延量は小さい（大きい）と言える。

[0029] 一方で、会話における発話のタイミングは、ユーザ各々の性格及び嗜好に応じて千差万別である。また同一ユーザであっても、その時点の身体的若しくは精神的負荷状態又は時間的余裕、或いはその他各種個別具体的な事情に応じて如何様にも変化し得る。会話特徴量取得手段によれば、このような多様に変化し得る発話タイミングを絶えず高精度に把握することができる。

[0030] このようにユーザの状態に即した会話特徴量に基づいてシステム総処理遅延量が決定されると、処理遅延量制御手段により、この決定されたシステム総処理遅延量が満たされるように会話データの処理遅延量が制御される。尚、処理遅延量制御手段に係る作用には、各拠点で負担すべき処理遅延量の決

定プロセスも含まれてよい。

[0031] ここで、実施形態に係るシステム総処理遅延量は、発話側の音声送受信装置と受話側の音声送受信装置からなるシステム全体として確保すべき処理遅延量の総量である。理論的には、システム総処理遅延量が増えない限り、会話のリズムやテンポ等が増えることはないから、快適性に係る会話品質は増えない。この点に鑑みれば、当該システム総処理遅延量が増えない限りにおいて、発話側及び受話側各々に対する処理遅延量の分配態様（即ち、各々における処理遅延の負担量である）には相応の自由度が生じる場合もある。このような場合については、システム総処理遅延量に対し、処理遅延量は多義的となり得る。

[0032] 但し、システム総処理遅延量を決定値に維持するためには、他拠点に設置される音声送受信システムとの協調が不可欠となる。この種の協調は、決定されたシステム総処理遅延量と自拠点側の処理遅延量の制御値とによって必然的に定まる要求値を、ネットワークを介して他拠点側の装置に告知することによってなされてもよい。或いは、決定されたシステム総処理遅延量を他拠点側の装置と共有し、予め策定されたアルゴリズムに従って両者協議の上で分配比率を決定すること等によってなされてもよい。例えば、各拠点における処理遅延量は、システム総処理遅延量を等分したものであってもよい。

[0033] このように、実施形態に係る音声送受信装置によれば、その時々ユーザの状態に即して、常に会話の音声品質と快適性との調和を図ることができる。即ち、最適な会話品質が提供されるのである。

[0034] 補足すると、予め実験的に、経験的に、理論的に又はシミュレーション等に基づいて処理遅延量或いはシステム総処理遅延量が決定されている場合、その時点におけるユーザの事情は、実は全く考慮されないに等しいから、事前に推定し得るユーザの基本的な性格等には一定の対応が可能となり得るものの、不確定要素や外乱要素に対する適応性は殆ど無いに等しい。従って、システム総処理遅延量は、ユーザ毎に或いは同一ユーザであってもその時々に応じて増減する真の最適値から乖離し易い。その結果、システム総処理遅

延量をより多くして音声品質を向上させ得るにもかかわらず不要にシステム総処理遅延量が抑制される、或いは逆にシステム総処理遅延量が大き過ぎて会話のテンポやリズムがユーザ固有のテンポやリズムから外れ、ユーザの快適性が低下する等の事態が、決して低くない頻度で発生してしまうのである。

[0035] また、会話に参加するユーザ同士で性格、嗜好及び各種事情が異なることも珍しくない。このような場合においても、実施形態に係る音声送受信装置によれば、自拠点ユーザ及び他拠点ユーザの各々について取得される会話特徴量に基づいてシステム総処理遅延量が設定されることにより、双方に不快感を生じさせない範囲で最大のシステム総処理遅延量を設定することができるため、実践上極めて有益である。

[0036] 尚、会話特徴量取得手段に係る「取得」とは、最終的に制御上の参照値として確定させることを意味しており、そのプロセスは何ら限定されない趣旨である。即ち、会話特徴量取得手段は、ネットワークを介する等して外部から会話特徴量を取得してもよいし、内部処理として算出、導出、推定、同定又は選択等の各種措置を講じることによって会話特徴量を取得してもよい。また、自拠点ユーザの会話特徴量取得プロセスと他拠点ユーザの会話特徴量取得プロセスとが相違していてもよい。

[0037] また、会話特徴量取得手段は、望ましくは、発話行為が発生する毎に、或いは一定又は不定の周期で、会話特徴量の取得を繰り返す。この場合、ユーザ側の最新の事情に対応し得るためより効果的である。また、過去に得られた会話特徴量に基づいた統計処理等が講じられる場合、会話特徴量の急変化が防止され、また会話特徴量の推定精度が向上するため効果的である。

[0038] 尚、実施形態に係る音声送受信装置に備わる、会話特徴量取得手段、処理遅延量決定手段及び処理遅延量制御手段は、夫々が或いは全体として、例えばCPU (Central Processing Unit) 又はMPU (Micro Processing Unit) 等の各種演算処理装置、各種プロセッサ、コントローラ又は各種機能モジュール等の各種形態を採り得る。

- [0039] 本発明の音声送受信装置に係る実施形態の一の態様では、前記会話特徴量取得手段は、前記会話特徴量として、前記自拠点において前記他拠点ユーザの発話出力が終了した時点から前記自拠点ユーザが発話を開始する時点までの時間である自拠点ユーザ反応時間と、前記他拠点において前記自拠点ユーザの発話出力が終了した時点から前記他拠点ユーザが発話を開始する時点までの時間である他拠点ユーザ反応時間とのうち少なくとも一方を取得する。
- [0040] この態様によれば、会話特徴量として、自拠点ユーザ反応時間又は他拠点ユーザ反応時間或いはその両方が取得される。
- [0041] ここで、「反応時間」とは、一方の発話出力が終了した時点から他方が発話を開始する時点までの時間であり、先述した「相手の発話終了後に発話を開始するタイミング」に相関する会話特徴量である。尚、「発話出力」とは、例えばスピーカ等の出力手段を介した出力を意味し、その開始及び終了時点は、相手の実際の発話行為の開始及び終了時点よりも時系列上後となる。即ち、反応時間とは、上述した処理遅延やその他各種遅延の影響を排除した、ユーザ各々の純粋な反応時間であり、個々のユーザにおける、会話の基本的なテンポ及びリズムを規定する時間である。従って、システム総処理遅延量の最適値を決定する上での参照値として好適である。
- [0042] このように、反応時間は、ユーザ固有の時間値であるから、自拠点ユーザ反応時間と他拠点ユーザ反応時間とは、当然ながら異なる方が自然であるが、これらが殆ど等しかろうが大きく異なろうが、これら自拠点ユーザ反応時間と他拠点ユーザ反応時間とのうち少なくとも一方に基づいて、より望ましくは両方に基づいてシステム総処理遅延量が決定されることにより、少なくとも必要量以上の快適性が担保された最適な会話品質を提供することができる。
- [0043] 他拠点ユーザ反応時間と自拠点ユーザ反応時間とが取得される本発明の音声送受信装置に係る実施形態の一の態様では、前記処理遅延量決定手段は、前記取得された他拠点ユーザ反応時間の大小に応じて夫々大小に変化する自拠点第1許容処理遅延量及び前記取得された自拠点ユーザ反応時間の大小に

応じて夫々大小に変化する他拠点第 1 許容処理遅延量のうち最小となる値以下の範囲で前記総量を決定する。

[0044] この態様によれば、自拠点第 1 許容処理遅延量及び他拠点第 1 許容処理遅延量のうち最小となる値（二者間の会話であれば、即ち小さい方の値）以下の範囲でシステム総処理遅延量が決定される。従って、自拠点ユーザ及び他拠点ユーザのいずれにおいても不快感の生じない会話品質を提供することができる。

[0045] 尚、自拠点第 1 許容処理遅延量及び他拠点第 1 許容処理遅延量の算出プロセスは、音声送受信システムを構築する如何なる構成要素においてなされてもよい。即ち、最終的にこれらの最小値以下の範囲でシステム総処理遅延量が決定される限りにおいて、これらの全てが実施形態に係る音声送受信装置でなされる必要はない。例えば、実施形態に係る音声送受信装置においては、ネットワークを介して他拠点で検出された他拠点ユーザ反応時間を取得し、自拠点第 1 許容処理遅延量を算出するのみであってもよい。この場合、他拠点においては、自拠点から送信される自拠点ユーザ反応時間が取得され、他拠点第 1 許容処理遅延量が算出されることになる。このように各装置で処理が分散されると、一の音声送受信装置に負荷が集中する事態を防ぐことができる。

[0046] 他拠点ユーザ反応時間と自拠点ユーザ反応時間とが取得される本発明の音声送受信装置に係る実施形態の他の態様では、前記通信手段は、前記他拠点に設置される音声送受信装置に対し、前記ネットワークを介して、前記取得された自拠点ユーザ反応時間に対応する自拠点ユーザ反応時間データを送信する。

[0047] この態様によれば、取得された自拠点ユーザ反応時間が、自拠点ユーザ反応時間データとして他拠点側へ送信される。従って、最終的にシステム総処理遅延量の決定に供される各種の参照値の算出プロセスのうち、自拠点ユーザ反応時間に基づいた一部を他拠点側の装置に委ねることが可能となり、処理負担を分散することが可能となる。

- [0048] 他拠点ユーザ反応時間と自拠点ユーザ反応時間とが取得される本発明の音声送受信装置に係る実施形態の他の態様では、前記会話特徴量取得手段は、前記会話特徴量として、前記自拠点において前記自拠点ユーザが発話を終了した時点から再び前記自拠点ユーザが発話を開始する時点までの時間である自拠点ユーザ反応待機時間を更に取得する。
- [0049] この態様によれば、会話特徴量として、自拠点ユーザ反応待機時間が取得される。
- [0050] ここで、「反応待機時間」とは、一方の発話出力が終了した時点から一方が再び発話を開始する時点までの時間であり、先述した「自分の発話終了後に相手が無反応であると判断して再度発話を開始するタイミング」に相關する会話特徴量である。即ち、反応待機時間とは、ユーザ各々の性格、嗜好及び各種事情を反映した時間であり、個々のユーザにおける、会話の基本的なテンポ及びリズムを規定する時間である。従って、システム総処理遅延量の最適値を決定する上での参照値として好適である。
- [0051] 尚、この態様では、前記取得された自拠点ユーザ反応待機時間と前記取得された他拠点ユーザ反応時間との差の大小に応じて夫々大小に変化する自拠点第2許容処理遅延量及び前記他拠点において前記他拠点ユーザが発話を終了した時点から再び前記他拠点ユーザが発話を開始する時点までの時間として取得された他拠点ユーザ反応待機時間と前記取得された自拠点ユーザ反応時間との差の大小に応じて夫々大小に変化する他拠点第2許容処理遅延量のうち最小となる値以下の範囲で前記総量を決定してもよい。
- [0052] この態様によれば、自拠点第2許容処理遅延量及び他拠点第2許容処理遅延量のうち最小となる値（二者間の会話であれば、即ち小さい方の値）以下の範囲でシステム総処理遅延量が決定される。従って、自拠点ユーザ及び他拠点ユーザのいずれにおいても不快感の生じない会話品質を提供することができる。
- [0053] ここで特に、反応待機時間と反応時間との差の大小に応じて夫々大小に変化する第2許容処理遅延量は、一方のユーザが、一方のユーザの発話に対し

他方のユーザが無反応であるとの誤判断の下に再度発話することによって生じる双方発話状態を回避する上で実践上極めて有益な参照値となり得る。

[0054] 尚、自拠点第2許容処理遅延量及び他拠点第2許容処理遅延量の算出プロセスは、音声送受信システムを構築する如何なる構成要素においてなされてもよい。即ち、最終的にこれらの最小値以下の範囲でシステム総処理遅延量が決定される限りにおいて、これらの全てが実施形態に係る音声送受信装置でなされる必要はない。例えば、実施形態に係る音声送受信装置においては、ネットワークを介して他拠点で検出された他拠点ユーザ反応時間を取得し、自拠点で取得された自拠点ユーザ反応待機時間を使用して自拠点第2許容処理遅延量を算出するのみであってもよい。この場合、他拠点においては、自拠点から送信される自拠点ユーザ反応時間が取得され、他拠点において取得された他拠点ユーザ反応待機時間を使用して他拠点第2許容処理遅延量が算出される。このように各装置で処理が分散されると、一の音声送受信装置に負荷が集中する事態を防ぐことができる。

[0055] 更に、この場合、前記処理遅延量決定手段は、前記取得された他拠点ユーザ反応時間の大小に応じて夫々大小に変化する自拠点第1許容処理遅延量及び前記取得された自拠点ユーザ反応時間の大小に応じて夫々大小に変化する他拠点第1許容処理遅延量並びに前記自拠点第2許容処理遅延量及び前記他拠点第2許容処理遅延量のうち最小となる値以下の範囲で前記総量を決定してもよい。

[0056] この態様によれば、夫々がユーザの状態を反映した参照値として算出された、自拠点第1許容処理遅延量、他拠点第1許容処理遅延量、自拠点第2許容処理遅延量及び他拠点第2許容処理遅延量のうち最小値以下の範囲でシステム総処理遅延量が決定される。従って、快適性の低下に伴う会話品質の低下をより確実に抑制することができる。

[0057] 本発明の音声送受信装置に係る実施形態の他の態様では、前記音声送受信装置は、前記会話データを前記送受信に相前後して一時的に蓄積するバッファを具備し、前記処理遅延量制御手段は、前記決定された総量に基づいて、

前記バッファに係るバッファ容量を制御する。

[0058] データパケット間に生じるジッタを吸収し、音切れやデータ欠落等による音声品質の低下を防止しようとする場合、バッファを設けるのが好適である。一方、バッファ容量の大小は、会話の時間遅延量の大小と一対一の関係にあり、無条件に大きくすることは全体的な会話品質上許容されない。即ち、バッファ容量は、実施形態に係る処理遅延量を制御するにあたっての処理遅延量制御手段の実制御対象として妥当である。

[0059] 尚、制御対象としてバッファ容量を捉える場合、決定されたシステム総処理遅延量に基づいて、音声送受信システムに備わるバッファ容量を如何に制御するかについては自由度がある。例えば、各音声送受信装置において、バッファが、受信データを受信後に一時的に蓄積する受信バッファと、送信データを送信前に一時的に蓄積する送信バッファとを含んで構築される場合、処理遅延量制御手段は、自拠点側の受信バッファ及び送信バッファと、他拠点側の受信バッファ及び送信バッファのバッファ容量の総和に対応する処理遅延量がシステム総処理遅延量となるように、比較的自由にこれら相互間のバッファ容量の分配比率を決定することができる。

[0060] 本発明の音声送受信装置に係る実施形態の他の態様では、前記処理遅延量制御手段は、前記決定された総量に基づいて、前記会話データを符号化するにあたっての符号化レートを制御する。

[0061] 遠隔地のユーザ同士で会話を成立させる場合、好適にはエンコード等の符号化プロセスが必要となる。

[0062] この際、自拠点での入力音声をエンコードするにあたってのエンコードビットレート等を意味する符号化レートの高低は、夫々音声品質の高低に対応し、且つ時間遅延量の大小に対応する。

[0063] 従って、この種の符号化レートは、実施形態に係る処理遅延量を制御するにあたっての処理遅延量制御手段の実制御対象として妥当である。

[0064] 本発明の音声送受信装置に係る実施形態の他の態様では、前記ネットワークの伝送状態を取得する伝送状態取得手段を更に具備し、前記処理遅延量決

定手段は、前記取得された会話特徴量と前記取得された伝送状態とに基づいて前記総量を決定する。

[0065] この態様によれば、伝送状態取得手段により取得されるネットワークの伝送状態が、システム総処理遅延量の決定に反映されるため、システム総処理遅延量をより正確に決定することができる。

[0066] 尚、この態様では、前記伝送状態取得手段は、前記ネットワークの伝送状態として、前記ネットワークの伝送遅延量を取得してもよい。

[0067] ネットワークの伝送遅延量は、即ち、処理遅延量と等しい次元で会話の時間遅延を規定する。従って、システム総処理遅延量の決定に反映させるべき伝送状態として好適である。尚、このようなネットワークの伝送遅延量とは、好適には、R T T (Round Trip Time) を指す。

[0068] 本発明の音声送受信装置に係る実施形態の他の態様では、前記取得された会話特徴量を統計処理する統計処理手段を更に具備し、前記処理遅延量決定手段は、前記統計処理された会話特徴量に基づいて前記総量を決定する。

[0069] この態様によれば、統計処理手段により、取得された会話特徴量に統計処理が施される。即ち、過去一定又は不定の期間にわたって取得された会話特徴量が、システム総処理遅延量の決定に反映させるべき会話特徴量に反映される。このため、会話特徴量の信頼度を向上させることができ、会話品質を安定的に維持することが可能となる。

[0070] 尚、統計処理の実践的態様は、特に限定されないが、好適な一形態として、過去一定期間にわたって取得された会話特徴量を加算平均する処理であってもよい。この際、明らかな異常値はサンプルから除外する等の措置が講じられてもよい。

[0071] 本発明の音声送受信装置に係る実施形態の他の態様では、前記取得された会話特徴量を記憶する記憶手段を更に具備する。

[0072] この態様によれば、取得された会話特徴量が、例えばH D D (Hard Disk Drive)、フラッシュメモリ、F D D (Floppy (登録商標) Disc Drive)、D V D或いはB D D (Blu-ray Disc Drive)等の各種態様を有し得る記憶装置

に揮発的に又は不揮発的に記憶されるため、システム総処理遅延量の決定を円滑に行うことができる。また、次回、同様に会話を行う際のシステム総処理遅延量の初期値を、この記憶された会話特徴量に基づいて決定することもできるため、システム総処理遅延量が最適なシステム総処理遅延量に収束するまでの時間を短縮化することも可能となる。

[0073] 本発明の音声送受信システムに係る実施形態は、各々がネットワークに收容される、複数の拠点に設置された音声送受信装置を含み、該音声送受信装置を介して、前記複数の拠点に夫々存在するユーザ相互間で会話を成立させることが可能な音声送受信システムであって、前記音声送受信装置は、前記複数の拠点のうち自拠点を除く他拠点に設置される前記音声送受信装置との間で、前記ネットワークを介して、少なくとも音声データを含む、前記会話の内容を表す会話データの送受信を行う通信手段と、前記自拠点到存在する自拠点ユーザ及び前記他拠点到存在する他拠点ユーザの各々について、前記会話における発話のタイミングに関連する所定の会話特徴量を取得する会話特徴量取得手段と、前記取得された会話特徴量に基づいて、その大小が前記会話の時間遅延量の大小に夫々対応し且つ前記会話の音声品質の高低に夫々対応する前記会話データの処理遅延量の、前記音声送受信システム全体における総量を決定する処理遅延量決定手段と、前記決定された総量が満たされるように前記処理遅延量を制御する処理遅延量制御手段とを具備する。

[0074] 音声送受信システムに係る実施形態は、上述した実施形態に係る音声送受信装置を備えるため、最適な会話品質を得ることが可能である。

[0075] 本発明のサーバ装置に係る実施形態は、各々がネットワークに收容される、サーバ装置と、複数の拠点に設置された、該複数の拠点のうち自拠点を除く他拠点に設置される音声送受信装置との間で、前記ネットワークを介して、少なくとも音声データを含む、前記会話の内容を表す会話データの送受信を行う通信手段を具備する音声送受信装置とを含み、該音声送受信装置を介して、前記複数の拠点に夫々存在するユーザ相互間で会話を成立させることが可能な音声送受信システムにおける、前記サーバ装置であって、前記複数

の音声送受信装置から、前記ネットワークを介して、前記自拠点に存在する自拠点ユーザ及び前記他拠点に存在する他拠点ユーザの各々についての前記会話における発話のタイミングに関連する所定の会話特徴量を取得する会話特徴量取得手段と、前記取得された会話特徴量に基づいて、その大小が前記会話の時間遅延量の大小に夫々対応し且つ前記会話の音声品質の高低に夫々対応する前記会話データの処理遅延量の、前記音声送受信システム全体における総量を決定する処理遅延量決定手段と、前記複数の音声送受信装置に対し前記ネットワークを介して前記決定された総量を告知する告知手段とを具備する。

[0076] サーバ装置に係る実施形態は、上述した音声送受信装置の実施形態に係る会話特徴量取得手段及び処理遅延量決定手段を備えるため、最適な会話品質を得ることが可能である。また、このように、会話特徴量の取得プロセス及びシステム総処理遅延量の決定プロセスをサーバ装置が担うことにより、音声送受信システムを構築する音声送受信装置の負担を著しく軽減することができ実践上非常に有益である。例えば、この際、各音声送受信装置では、サーバ装置に備わる告知手段により告知されたシステム総処理遅延量の最適値が満たされるように、例えば処理遅延量制御手段に類する手段が処理遅延量を制御すればよい。或いは、サーバ装置側で然るべきコマンドを実行して、音声送受信装置を上位制御してもよい。また、サーバ装置側で、システム総処理遅延量に基づいた各音声送受信装置の処理遅延量の分配態様をも規定し得る場合には、音声送受信装置側の負担を一層軽減することも可能である。

[0077] 以上説明したように、本発明の音声送受信装置に係る実施形態によれば、通信手段、会話特徴量取得手段、処理遅延量決定手段及び処理遅延量制御手段を備えるので、最適な会話品質を得ることができる。

[0078] 以上説明したように、本発明の音声送受信システムに係る実施形態によれば、本発明の音声送受信装置に係る実施形態を備えるので、最適な会話品質を得ることができる。

[0079] 以上説明したように、本発明の音声送受信装置に係る実施形態によれば、

会話特微量取得手段、処理遅延量決定手段及び告知手段を備えるので、最適な会話品質を得ることができる。

[0080] 本発明のこのような作用及び他の利得は次に説明する実施例から明らかにされる。

実施例

[0081] 以下、適宜図面を参照して、本発明の好適な各種実施例について説明する。

<第1実施例>

<実施例の構成>

始めに、図1を参照し、本発明の第1実施例に係る遠隔会議システム1の構成について説明する。ここに、図1は、遠隔会議システム1の構成を概念的に表してなる概略構成図である。

[0082] 図1において、遠隔会議システム1は、相互いに離れた拠点X（即ち、本発明に係る「自拠点」の一例）及び拠点Y（即ち、本発明に係る「他拠点」の一例）を繋ぐ広域ネットワーク20（IP（Internet Protocol）網）に収容された、本発明に係る「音声送受信システム」の一例たる音声会議システムである。

[0083] 遠隔会議システム1は、拠点Xに設置され、拠点Xに存在するユーザA（即ち、本発明に係る「自拠点ユーザ」の一例である）に使用される音声送受信装置10A（即ち、本発明に係る「音声送受信装置」の一例）と、拠点Yに設置され、拠点Yに存在するユーザB（即ち、本発明に係る「他拠点ユーザ」の一例である）に使用される音声送受信装置10B（即ち、本発明に係る「音声送受信装置」の他の一例）から構成される。ユーザAとユーザBとは、遠隔会議システム1により音声情報のやり取りを介した円滑な音声会議を行うことができる。

[0084] 次に、図2を参照し、遠隔会議システム1を構成する一の音声送受信装置の構成について、その動作を交えて説明する。ここに、図2は、音声送受信装置10Aの構成を概念的に表してなるブロック図である。尚、同図におい

て、図 1 と重複する箇所には、同一の符号を付してその説明を適宜省略することとする。

[0085] 尚、遠隔会議システム 1 においては、音声送受信装置 10A のハードウェア構成は、音声送受信装置 10B と同等であるとする。

[0086] 図 2 において、音声送受信装置 10A は、音声入力ユニット 100、音声出力ユニット 200、会話特徴量検出部 300、会話特徴量統計処理部 400、記憶装置 500、処理遅延量決定部 600、RTT 測定部 700、処理遅延情報通信部 800 及びバッファ制御部 900 を備える。

[0087] 音声入力ユニット 100 は、音声入力部 110、エンコーダ 120、送信バッファ 130 及び音声データ送信部 140 を備え、ユーザ 10A の発話音声 を音声データとしてネットワーク 20 を介して音声送受信装置 10B へ送信可能なユニットである。

[0088] 音声入力部 110 は、入力端子（符合省略）が図示せぬマイクに接続された入力インターフェイスであり、当該マイクを介して入力されたユーザ A の発話音声をアナログ音声信号として取り込み可能に構成されている。

[0089] エンコーダ 120 は、音声入力部 110 を介して入力されるアナログ音声信号を、所定の符号化レート（エンコードビットレート）で符号化し、デジタル音声データに変換するデジタル変換装置である。エンコーダ 120 に係るアナログ音声データのエンコード態様としては、公知の各種規格に準じたものを採用可能である。例えば、この種の規格とは、MPEG（Moving Picture Expert Group）等の規格であってもよい。

[0090] 送信バッファ 130 は、エンコーダ 120 を介して得られたデジタル音声データを所定の送信バッファ容量 D_{a_s} に相当するデータ量だけ一時的に蓄積する揮発性記憶装置である。

[0091] 音声データ送信部 140 は、出力端子（符合省略）がネットワーク 20 に接続された送信インターフェイスであり、送信バッファ 130 から順次出力されてくるデジタル音声データを順次ネットワーク 20 を介して音声入出力装置 10B に送信可能に構成される。即ち、音声データ送信部 140 は、本

発明に係る「通信部」の一例であり、送信されるデジタル音声データは、本発明に係る「会話データ」の一例である。

[0092] 音声出力ユニット200は、音声データ受信部210、受信バッファ220、デコーダ230及び音声出力部240を備え、ユーザBの発話音声スピーカ等の出力装置を介して出力可能なユニットである。

[0093] 音声データ受信部210は、出力端子（符合省略）がネットワーク20に接続された受信インターフェイスであり、ネットワーク20を介して送信されてくるユーザBの発話音声に対応するデジタル音声データを順次取り込み可能に構成される。即ち、音声データ受信部210は、本発明に係る「通信部」の他の一例であり、受信されるデジタル音声データは、本発明に係る「会話データ」の他の一例である。

[0094] 受信バッファ220は、音声データ受信部210を介して得られたデジタル音声データを所定の受信バッファ容量D_aに相当するデータ量だけ一時的に蓄積する揮発性記憶装置である。

[0095] デコーダ230は、受信バッファ220から順次出力されてくるデジタル音声データを復号化しアナログ音声データに変換するアナログ変換装置である。デコーダ230の機能は、エンコーダ120のエンコード機能と対をなすものであり、両者は当然ながら同一の規格に準じたデータ変換を行うように構成される。

[0096] 音声出力部240は、出力端子（符合省略）が図示せぬスピーカに接続された出力インターフェイスであり、当該スピーカを介してユーザBの発話音声出力可能に構成されている。

[0097] 会話特徴量検出部300は、ユーザAの会話特徴量として後述するユーザA反応時間R_A及びユーザA反応待機時間T_Aを検出可能に構成された、本発明に係る「会話特徴量取得手段」の一例である。

[0098] 会話特徴量統計処理部400は、会話特徴量検出部300により適宜検出される会話特徴量を統計処理可能に構成された、本発明に係る「統計処理手段」の一例である。会話特徴量統計処理部400は、検出された会話特徴量

を過去一定サンプルについて保持可能な構成となっており、保持するサンプルのサンプル値を加算平均処理して出力する構成となっている。

[0099] 記憶装置 500 は、例えば HDD やフラッシュメモリ等の不揮発性記憶装置であり、本発明に係る「記憶手段」の一例である。記憶装置 500 の記憶領域には、会話特徴量統計処理部 400 を介して出力されたユーザ A 反応時間 R_A の平均値を表す反応時間データ DAT_R_A 及びユーザ A 反応待機時間 T_A の平均値を表す反応待機時間データ DAT_T_A が格納される構成となっている。

[0100] 処理遅延量決定部 600 は、許容処理遅延推定部 610 及び交渉部 620 を備え、送信バッファ 130 のバッファ容量 D_{a_s} 及び受信バッファ 220 のバッファ容量 D_{a_r} を決定するように構成された、本発明に係る「処理遅延量決定手段」の一例たる制御装置である。処理遅延量決定部 600 は、ROM (Read Only Memory) に格納された制御プログラムに従って、後述するバッファ容量制御を実行可能に構成されている。

[0101] 許容処理遅延推定部 610 は、後述する、遠隔会議システム 1 に許容される最大の処理遅延量たる最大許容処理遅延量 d_{max} (即ち、本発明に係る「会話データの処理遅延量の音声送受信システム全体における総量」の一例である) を決定するための各種処理を実行するプロセッサである。許容処理遅延推定部 610 は、予め理論的に構築された計算モデルとしての第 1 計算モデル及び第 2 計算モデルを備え、これら計算モデルに基づいて当該処理を実行する構成となっている。

[0102] 交渉部 620 は、拠点 Y の音声送受信装置 10B との間で、本発明に係る「処理遅延量」の一例としてのバッファ容量の分配交渉を行うプロセッサである。交渉部 620 はまた、この音声送受信装置 10B との交渉を経て、送信バッファ容量 D_{a_s} 及び受信バッファ容量 D_{a_r} の制御目標値を決定する。

[0103] RTT 測定部 700 は、ネットワーク 20 の伝送遅延量である RTT を測定可能に構成されたプロセッサである。RTT 測定部 700 は、RTCP (R

real-time Transport Control Protocol) の S R (Sender Report) や R R (Receiver Report) を利用して R T T を測定する。

[0104] 尚、ここで規定される R T T は、音声送受信装置 10 A において音声データ送信部 140 を介してデジタル音声データが送信された時点から音声送受信装置 10 B において音声データ受信部 210 を介して当該デジタル音声データが受信される時点までの時間と、音声送受信装置 10 B において音声データ送信部 140 を介してデジタル音声データが送信された時点から音声送受信装置 10 A において音声データ受信部 210 を介して当該デジタル音声データが受信される時点までの時間との和である。ユーザ A とユーザ B とが同一拠点で対面状態で会話を行っている場合等、通常の会話においては、この R T T に相当する時間遅延は存在しない。即ち、遠隔会議システム 1 において、この R T T は、遠隔会議システム 1 を利用したユーザ相互間の会話の時間遅延量を規定する指標値となる。

[0105] 処理遅延情報通信部 800 は、出力端子がネットワーク 20 に接続されており、音声送受信装置 10 B との間で、上述したバッファ量の目標値策定に係る各種のデータの送受信を行う通信インターフェイスである。処理遅延情報通信部 800 は、音声送受信装置 10 B に対し、ユーザ A 反応時間 R_A 及び各計算モデルに基づいて設定された処理遅延量 d を送信し、音声送受信装置 10 B からユーザ B 反応時間 R_B 及び各計算モデルに基づいて設定された処理遅延量 d を取得する。

[0106] バッファ制御部 900 は、送信バッファ 130 及び受信バッファ 220 について、各バッファ容量を可変に制御可能に構成された、本発明に係る「処理遅延量制御手段」の一例たるプロセッサである。

<実施例の動作>

次に、図 3 を参照し、本実施例の動作として、処理遅延量決定部 600 により実行されるバッファ容量制御の詳細について説明する。ここに、図 3 は、バッファ容量制御のフローチャートである。尚、音声送受信装置 10 A では、このバッファ容量制御とは別に、音声入力ユニット 100 及び音声出力

ユニット 200 による先に述べた会話データの送受信が適宜実行されている。

- [0107] 図 3 において、バッファ容量制御では、先ずユーザ A 反応時間 R_A の送信処理及びユーザ B 反応時間 R_B の受信処理が実行される（ステップ S 101）。ユーザ A 反応時間 R_A 及びユーザ B 反応時間 R_B は、夫々、本発明に係る「会話特徴量」の一例であり、また前者は「自拠点ユーザ反応時間」、後者は「他拠点ユーザ反応時間」の夫々一例である。
- [0108] ステップ S 101 においては、記憶装置 500 から反応時間データ DAT_R_A が読み出され、処理遅延情報通信部 800 によりネットワーク 20 を介して音声送受信装置 10B へ送信される。一方で、処理遅延情報通信部 800 によりネットワーク 20 を介して音声送受信装置 10B からユーザ B 反応時間 R_B に対応する反応時間データ DAT_R_B が取得され、許容処理遅延推定部 610 へ送出される。
- [0109] ここで、図 4 を参照し、反応時間について説明する。ここに、図 4 は、反応時間の概念を説明するタイミングチャートである。
- [0110] 図 4 において、ユーザ A とユーザ B とが、互いの音声伝達に有意な時間遅延が生じない理想的な環境で会話しているとする。尚、このような理想的な環境とは、例えば、同一拠点において、対面状態で会話がなされる環境等を指す。
- [0111] ある任意の時刻 T_0 において、ユーザ A の発話が終了したとする（ユーザ A のハッチング部分参照）。一方、ユーザ B は、時刻 T_0 においてユーザ A の発話終了を知覚する。そして、この発話終了時点（ T_0 ）から、ユーザ B に固有の遅延時間を経た時刻 T_1 において発話を開始する（ユーザ B のハッチング部分参照）。この一方の発話終了時点から他方の発話開始時点までの遅延時間が、反応時間である。反応時間は、ユーザの性格、嗜好、精神的負荷状態、肉体的負荷状態、及びその他その都度個別具体的に変化し得る各種の事情に応じて多様に変化する。
- [0112] ユーザ A 反応時間 R_A とは、拠点 X におけるユーザ A の反応時間であり、ユ

ーザB反応時間 R_B とは、拠点YにおけるユーザBの反応時間である。尚、遠隔会議システム1のように遠隔地同士の会話である場合、一方の発話終了時点とは、スピーカ等の音声出力手段を介した音声出力（即ち、本発明に係る「発話出力」の一例である）の終了時点の意味する。一方の発話行為が終了したところで発話内容が他方に認識されなければ、会話が成立しないからである。

[0113] 反応時間は、夫々の音声送受信装置において、会話特徴量検出部300により実行される反応時間算出処理により算出される。ここで、図5を参照し、反応時間算出処理について説明する。ここに、図5は、反応時間算出処理のフローチャートである。尚、図5の反応時間算出処理は、拠点Xにおいて実行される処理であるとする。

[0114] 図5において、先ずユーザBの音声出力が有るか否かが判別される（ステップS201）。ユーザBの音声出力が無い場合（ステップS201：NO）、処理はステップS201に戻され一連の処理が繰り返される。ユーザBの音声出力が有る場合（ステップS201：YES）、その時点の時刻が最終音声出力時刻 T_{op} として更新される（ステップS202）。

[0115] 最終音声出力時刻 T_{op} が更新されると、ユーザBの音声出力が無いか否かが判別される（ステップS203）。ユーザBの音声出力が継続している場合（ステップS203：NO）、処理はステップS202に戻され、最終音声出力時刻 T_{op} の更新が継続される。一方、ユーザBの音声出力が無い場合（ステップS203：YES）、ユーザAの音声入力があるか否かが判別される（ステップS204）。

[0116] 尚、ユーザBが必ずしも連続的に発話しているとは限らないため、ユーザBの音声出力が無い旨の判別は、ユーザBが一連の発話中であるにもかかわらず発話が終了したと誤判別されぬように、無音区間の長さが予め設定された基準値を超えたか否かに基づいて正確になされる構成となっている。

[0117] ユーザAの音声入力が無い場合（ステップS204：NO）、即ち、ユーザBの発話出力終了後、ユーザAがそれに対する発話内容を考えていると推

定される間は、ステップS 2 0 3が繰り返し実行される。ユーザAの音声入力が開始される（発話が開始される）と（ステップS 2 0 4：YES）、その時点の時刻TとステップS 2 0 2において更新された最終音声出力時刻T_{o p}との差に相当する時間値が、ユーザA反応時間R_Aとして決定される（ステップS 2 0 5）。ユーザA反応時間R_Aが算出されると、処理はステップS 2 0 1へ戻され、一連の処理が繰り返される。反応時間算出処理は以上のように実行される。

[0118] 尚、ユーザB反応時間R_Bも、同様に音声送受信装置10Bにおいて算出されている。

[0119] 反応時間算出処理により適宜算出されたユーザA反応時間R_Aは、算出される毎に会話特徴量統計処理部400に送出され、統計処理に供される。統計処理は先述したように過去所定サンプル分についての加算平均処理である。尚、統計処理の態様は、加算平均処理に限定されない。

[0120] 会話特徴量統計処理部400により統計処理を施されたユーザA反応時間R_Aは、記憶部500に反応時間データDAT__R_Aとして格納される。反応時間データDAT__R_Aは、ユーザA反応時間R_Aが算出され統計処理が実行される毎に適宜更新される。

[0121] 図3に戻り、反応時間データの送受信が終了すると、ユーザA反応待機時間T_Aが取得される（ステップS 1 0 2）。

[0122] 反応待機時間とは、一方のユーザの発話が終了した時点から、当該一方のユーザが再度発話を開始する時点までの時間値である。反応待機時間は、ユーザの性格、嗜好、精神的負荷状態、肉体的負荷状態、及びその他その都度個別具体的に变化し得る各種の事情に応じて多様に変化する。

[0123] 反応待機時間は、夫々の音声送受信装置において、会話特徴量検出部300により実行される反応待機時間算出処理により算出される。ここで、図6を参照し、反応待機時間算出処理について説明する。ここに、図6は、反応待機時間算出処理のフローチャートである。尚、図6の反応待機時間算出処理は、拠点Xにおいて実行される処理であるとする。

- [0124] 図6において、先ずユーザAの音声入力があるか否かが判別される（ステップS301）。ユーザAの音声入力がない場合（ステップS301：NO）、処理はステップS201に戻され一連の処理が繰り返される。ユーザAの音声入力がある場合（ステップS201：YES）、その時点の時刻が最終音声入力時刻 T_{ip} として更新される（ステップS302）。
- [0125] 最終音声入力時刻 T_{ip} が更新されると、ユーザBの音声出力がないか否かが判別される（ステップS303）。ユーザBの音声出力がある場合（ステップS303：NO）、ユーザA反応待機時間 T_A よりも短い時間でユーザBの応答が返ってきたものとして、処理はステップS301に戻る。
- [0126] ユーザBの音声出力がない場合（ステップS303：YES）、ユーザAの音声入力があるか否かが判別される（ステップS304）。ユーザAの音声入力がない場合（ステップS304：NO）、処理はステップS303に戻る。即ち、ユーザAの音声入力が開されるまで、処理は待機状態となる。
- [0127] ユーザAの音声入力がある場合（ステップS304：YES）、現時点の時刻 T とステップS302で更新された最終音声入力時刻 T_{ip} との差に相当する時間値が、基準値 T_0 よりも大きいか否かが判別される。ここで、ユーザAは必ずしも連続的に発話している訳ではないため、一連の発話中であっても一時的に発話が途切れることがある。基準値 T_0 は、ユーザAの音声入力が、このような一連の発話動作に相当するか否かを判別するための判断基準値である。
- [0128] 即ち、 $T - T_{ip}$ に相当する時間値が基準値 T_0 以下である場合（ステップS305：NO）、処理はステップS301に戻る。一方、 $T - T_{ip}$ に相当する時間値が基準値 T_0 よりも大きい場合（ステップS305：YES）、 $T - T_{ip}$ に相当する時間値がユーザA反応待機時間 T_A として決定される（ステップS306）。反応待機時間算出処理は以上のように実行される。
- [0129] 尚、ユーザB反応待機時間 T_B も、同様に音声送受信装置10Bにおいて算

出されている。

[0130] 反応時間算出処理により適宜算出されたユーザA反応待機時間 T_A は、算出される毎に会話特徴量統計処理部400に送出され、統計処理に供される。統計処理は先述したように過去所定サンプル分についての加算平均処理である。尚、統計処理の態様は、加算平均処理に限定されない。

[0131] 会話特徴量統計処理部400により統計処理を施されたユーザA反応待機時間 T_A は、記憶部500に反応時間データ DAT_T_A として格納される。反応時間データ DAT_T_A は、ユーザA反応待機時間 T_A が算出され統計処理が実行される毎に適宜更新される。

[0132] 再び図3に戻り、ユーザA反応待機時間 T_A が取得されると、第1計算モデル及び第2計算モデルに基づいて処理遅延量 d が設定される（ステップS103）。処理遅延量 d は、第1計算モデル及び第2計算モデルの各々について設定される。尚、処理遅延量 d は、遠隔会議システム1の処理遅延量の総量を意味する。

[0133] ここで、図7を参照し、第1計算モデルについて説明する。ここに、図7は、第1計算モデルの概念を説明するタイミングチャートである。尚、同図において、図4と重複する箇所には、同一の符号を付してその説明を適宜省略することとする。

[0134] 図7において、遠隔会議システム1における処理遅延量 d を下記（1）式により定義する。

[0135] $d = d_a + d_b + d_{proc} \cdots (1)$

上記（1）式において、 d_a は音声送受信装置10Aのバッファ容量と一対一に対応する遅延量であり、受信バッファ容量 D_{a_r} に対応する受信バッファ遅延量 d_{a_r} 及び送信バッファ容量 D_{a_s} に対応する送信バッファ遅延量 d_{a_s} との間に「 $d_a = d_{a_r} + d_{a_s}$ 」なる関係を有する。

[0136] また、上記（1）式において、 d_b は音声送受信装置10Bのバッファ容量と一対一に対応する遅延量であり、受信バッファ容量 D_{b_r} に対応する

受信バッファ遅延量 $d b_r$ 及び送信バッファ容量 $D b_s$ に対応する送信バッファ遅延量 $d b_s$ との間に「 $d b = d b_r + d b_s$ 」なる関係を有する。

[0137] 更に、上記（１）式において、 $d p r o c$ は、音声送受信装置 10A 及び 10B の各々におけるエンコーダ及びデコーダの処理遅延である。

[0138] より具体的には、 $d p r o c$ は、音声送受信装置 10A におけるエンコーダの処理遅延 $d e n c a$ 及びデコーダの処理遅延 $d d e c a$ 並びに音声送受信装置 10B におけるエンコーダの処理遅延 $d e n c b$ 及びデコーダの処理遅延 $d d e c b$ の総和である。尚、 $d p r o c$ は、エンコーダ及びデコーダにおける符号化レートが一定であれば一定値となる。

[0139] 図 7 において、時刻 $T 0$ にユーザ A の発話が終了したとする。ここで、遠隔地同士である拠点 X と拠点 Y とにおいて遠隔会議システム 1 を利用して会話をを行う構成では、このユーザ A の発話終了が認識される時点が、図 4 のような通常会話時と異なる。

[0140] 即ち、拠点 Y においてユーザ A の発話終了が認識されるのは、時刻 $T 0$ から音声送受信装置 10A の送信バッファ遅延量 $d a_s$ 、音声送受信装置 10B の受信バッファ遅延量 $d b_r$ 、音声送受信装置 10A のエンコーダ処理遅延 $d e n c a$ 、音声送受信装置 10B のデコーダ処理遅延 $d d e c b$ 及び音声送受信装置 10A から音声送受信装置 10B へ向かうネットワーク 20 の片道遅延 $O W D a$ を経た後の時刻 $T 0'$ である。即ち、下記（２）式が成立する。

[0141]
$$T 0' = T 0 + d a_s + d b_r + d e n c a + d d e c b + O W D a$$

$$\dots (2)$$

尚、ネットワーク 20 の片道遅延 $O W D a$ に関しては、下記（３）式が成立する。

[0142]
$$R T T = O W D a + O W D b \dots (3)$$

尚、 $O W D b$ は、音声送受信装置 10B から音声送受信装置 10A へ向かうネットワーク 20 の片道遅延である。

[0143] 次に、時刻 T_0' からユーザB反応時間 R_B を経た時刻 T_1' においてユーザBが発話を開始したとする。このユーザBの発話開始が認識される時刻もまた、拠点Xと拠点Yとは異なり、拠点XにおいてユーザBの発話開始が認識されるのは、時刻 T_1' から音声送受信装置10Bの送信バッファ遅延量 d_{b_s} 、音声送受信装置10Aの受信バッファ遅延量 d_{a_r} 、音声送受信装置10Bのエンコーダ処理遅延 d_{encb} 、音声送受信装置10Aのデコーダ処理遅延 $d_{dec a}$ 及び上記片道遅延 OWD_b を経た後の時刻 T_2 となる。即ち、下記(4)式が成立する。

$$[0144] \quad T_2 = T_0' + R_B + d_{b_s} + d_{a_r} + d_{encb} + d_{dec a} + OW D_b \cdots (4)$$

即ち、図4に例示された通常会話時と同様の感覚で、ユーザAの発話終了後にユーザB反応時間 R_B を隔ててユーザBが発話を行ったとしても、遠隔会議システム1とネットワーク20の影響により、ユーザAにユーザBの発話開始が認識されるのは、下記(5)式に規定される遅延時間 T_L が経過した後となる。

$$[0145] \quad T_L = T_2 - T_0 = R_B + R_{TT} + d \cdots (5)$$

ここで、第1計算モデルでは、下記(6)式の指標値 Z を使用する。

$$[0146] \quad Z = T_L / R_B \cdots (6)$$

指標値 Z は、1より大きい値を採り、通常の会話環境(図4に例示される如き環境)に近づく程1に漸近する性質を持つ。また指標値 Z が大きくなる程、ユーザAは、会話が円滑に進まないと感じるようになり、快適性の低下を覚え得る。

[0147] 一方、この指標値 Z には予め実験的に最大値 F (例えば、 $F = 1.2$) が設定される。最大値 F は、それよりも大きい領域においてユーザAが感じる会話品質の低下が無視出来なくなると判断され得る適合値である。逆に言えば、指標値 Z が最大値 F 以下であれば、ユーザAはユーザBとの会話に実践上無視し得ない円滑性の欠如を感じることはない。

[0148] 具体的には、条件式として下記(7)式が設定される。

[0149] $(R_B + R_{TT} + d) / R_B \leq F \cdots (7)$

更に(7)式を変形すると、下記(8)式が得られる。

[0150] $d \leq (F - 1) R_B - R_{TT} \cdots (8)$

上記(8)式は、処理遅延量 d の採り得る範囲を規定する式である。尚、上記(8)式における処理遅延量 d は、本発明に係る「自拠点第1許容処理遅延量」の一例である。

[0151] 上記(8)式に示される処理遅延量 d は、ユーザAに対し設定される値であり、ユーザBに対しては、音声送受信装置10Bにおいて、同様のプロセスを経て、下記(9)式により処理遅延量 d が設定される。下記(9)式における処理遅延量 d は、本発明に係る「他拠点第1許容処理遅延量」の一例である。

[0152] $d \leq (F - 1) R_A - R_{TT} \cdots (9)$

次に、図8を参照し、第2計算モデルについて説明する。ここに、図8は、第2計算モデルの概念を説明するタイミングチャートである。尚、同図において、図7と重複する箇所には、同一の符号を付してその説明を適宜省略することとする。

[0153] 図8は、図7に対し、ユーザA反応待機時間 T_A の概念を表したものである。

[0154] ユーザA反応待機時間 T_A は、ユーザAの発話が終了した時刻 T_0 から、ユーザAが再び発話を開始する時刻 T_1' までの時間値である。ここで、計算モデル1の説明で述べたように、ユーザAの発話終了後、ユーザBの発話開始がユーザAによって認識されるのは、時刻 T_0 から上記遅延時間 T_L が経過した後の時刻 T_2 である。

[0155] ところが、この時刻 T_2 においては、既にユーザBの反応が無いものとしてユーザAが発話を開始しており、ユーザAとユーザBとが共に発話している状況が生じ得る。従って、拠点Xでの会話の円滑性を担保しようとするれば、このユーザA反応待機時間 T_A を考慮する必要がある。第2計算モデルは、このユーザA反応待機時間 T_A を考慮した計算モデルである。

[0156] 上述した二者同時発話の状況を回避するためには、上記遅延時間 T_L がユーザ A 反応待機時間 T_A 以下であればよいから、下記 (10) 式が成立する。

$$[0157] \quad R_B + R_{TT} + d \leq T_A \cdots (10)$$

上記 (10) 式を変形すると、下記 (11) 式が得られる。

$$[0158] \quad d \leq (T_A - R_B) - R_{TT} \cdots (11)$$

上記 (11) 式は、処理遅延量 d の採り得る範囲を規定する式である。尚、上記 (11) 式における処理遅延量 d は、本発明に係る「自拠点第 2 許容処理遅延量」の一例である。

[0159] 尚、上記 (11) 式に示される処理遅延量 d は、ユーザ A に対し設定される値であり、ユーザ B に対しては、音声送受信装置 10B において、同様のプロセスを経て、下記 (12) 式により処理遅延量 d が算出される。下記 (12) 式に示される処理遅延量 d は、本発明に係る「他拠点第 2 許容処理遅延量」の一例である。

$$[0160] \quad d \leq (T_B - R_A) - R_{TT} \cdots (12)$$

第 1 計算モデル及び第 2 計算モデルにより処理遅延量 d が設定されると、最大許容処理遅延量 d_{max} が決定される (ステップ S104)。最大許容処理遅延量 d_{max} は、ユーザ A 及びユーザ B の双方にとって、会話の円滑性が担保され得る処理遅延量である。従って、第 1 計算モデル及び第 2 計算モデルに基づいて設定された処理遅延量を比較する必要がある。より具体的には、最大許容処理遅延量 d_{max} は、上記 (8) 式、(9) 式、(11) 式及び (12) 式のいずれも満たす処理遅延量である必要がある。

[0161] そこで、ステップ S104 においては、処理遅延情報通信部 800 を介して、音声送受信装置 10B からユーザ B について第 1 及び第 2 計算モデルに基づいて設定された処理遅延量 d が取得される。一方で、処理遅延情報通信部 800 を介して、音声送受信装置 10B に対しユーザ A について第 1 及び第 2 計算モデルに基づいて設定された処理遅延量 d が送信される。即ち、音声送受信装置 10A と音声送受信装置 10B との間で、最大許容処理遅延量 d_{max} の決定に係る条件式が共有される。

- [0162] 続いて、上記（８）式、（９）式、（１１）式及び（１２）式のいずれも満たす処理遅延量 d が決定される。尚、具体的には、これら四式のうち、右辺項が最も小さい条件式によって規定される処理遅延量 d が、これら四式を満たす処理遅延量 d となる。
- [0163] 一方、音切れや符合誤り等による音声品質の低下を防止する観点からは、処理遅延量は大きい方が良い。従って、最終的に、最大許容処理遅延量 d_{max} は、これら四式を満たす範囲の最大値として決定される。
- [0164] 最大許容処理遅延量 d_{max} が決定されると、分配処理が実行される（ステップ $S105$ ）。分配処理とは、決定された最大許容処理遅延量 d_{max} を音声送受信装置 $10A$ 及び音声送受信装置 $10B$ の各バッファ容量に分配する処理である。この分配処理は、交渉部 620 により実行される。
- [0165] 具体的には、交渉部 620 は、処理遅延情報通信部 800 を介して音声送受信装置 $10B$ に負荷状況を確認する。音声送受信装置 $10B$ に処理負荷上の問題がなく、音声送受信装置 $10A$ にも処理負荷上の問題がなければ、交渉部 620 は、最大許容処理遅延量 d_{max} の 50% を音声送受信装置 $10A$ で負担する旨を決定し、またその旨を音声送受信装置 $10B$ に伝達する。その結果、処理負荷上の問題がなければ、通常、最大許容処理遅延量 d_{max} の 50% が音声送受信装置 $10A$ 側で負担される。
- [0166] 更に、交渉部 620 は、音声送受信装置 $10A$ で負担すべき遅延量を送信バッファ 130 及び受信バッファ 220 で分配する。通常、ここでも、送信バッファ 130 と受信バッファ 220 との負担率は等しく設定される。即ち、送信バッファ 130 に係る処理遅延量は、最大許容処理遅延量 d_{max} の 25% 、受信バッファ 220 に係る処理遅延量も、最大許容処理遅延量 d_{max} の 25% に設定される。設定された分配比率は、バッファ制御部 900 に伝達される。
- [0167] 分配比率を伝達されたバッファ制御部 900 は、設定された分配比率に応じた処理遅延量が得られるように、送信バッファ 130 及び受信バッファ 220 の容量を制御する（ステップ $S106$ ）。バッファ容量の制御がなされ

ると、処理はステップS 1 0 1に戻され、一連の処理が繰り返される。バッファ容量制御は以上のように実行される。

[0168] このように、本実施例に係るバッファ容量制御によれば、第1計算モデル及び第2計算モデルに基づいて設定された処理遅延量 d から最大許容処理遅延量 d_{max} が決定され、受信バッファ容量 D_{a_r} 及び送信バッファ容量 D_{a_s} の制御を介して、遠隔会議システム1の処理遅延量 d が、この最大許容処理遅延量 d_{max} に制御される。ここで、第1計算モデル及び第2計算モデルは、夫々ユーザの会話特徴量を反映し得るように構築されており、決定される最大許容処理遅延量 d_{max} は、会話の当事者たるユーザA及びユーザBの実情に即した、双方にとって円滑性が最低限担保されたものとなる。

[0169] 従って、会話の円滑性が阻害されることによる快適性の低下が防止される。一方で、最大許容処理遅延量 d_{max} は、この種の快適性の低下を防止し得る範囲で最大の値であり、音切れや符号化誤り等による音声品質の低下も防止される。即ち、本実施例によれば、会話に参加するユーザの事情に即した最適な会話品質が提供されるのである。

[0170] 尚、本実施例では、会話に参加するユーザは、ユーザAとユーザBとの二者となっているが、三者或いはより多くの参加者を有する会議や会話等においても、上記と同様の概念を適用してユーザの事情に即した最適な会話品質を提供することが可能であることは言うまでもない。

[0171] 尚、本実施例では、音声会議が前提とされたが、各拠点に撮像手段を配置して、音声データと共に画像データ或いは映像データの送受信を行うことも可能である。この場合、例えばTV会議等と称される会議形態を採ることも可能である。

[0172] 尚、本実施例に係る音声送受信装置は、記憶部500を備えるため、複数のユーザの過去の会話特徴量を保持しておくことも可能である。無論、過去の会話特徴量が次回においてそのまま運用できる保証はないが、例えば初期値として用いることによって、会話特徴量取得部300及び処理遅延量決定

部 600 の動作初期における、会話品質のばらつき等を抑制することができるため、好適である。

- [0173] 尚、本実施例では、バッファ容量の制御を介して処理遅延量の制御が実現されたが、処理遅延量と相関する制御量は、音声データの符号化レート（例えば、エンコーダ 120 のエンコードビットレート）であってもよい。符号化レートが高くなれば、相対的に高音質となる分、データ伝送量も増えるため、処理遅延量は大きくなる。従って、上述したバッファ容量に替えて或いは加えて符号化レートを制御することにより、上記と同様の効果を得ることが可能である。

<第 2 実施例>

次に、図 9 を参照し、本発明の第 2 実施例について説明する。ここに、図 9 は、本発明の第 2 実施例に係る遠隔会議システム 2 の構成を概念的に表してなる概略構成図である。尚、同図において、図 1 と重複する箇所には同一の符号を付してその説明を適宜省略することとする。

- [0174] 図 9 において、遠隔会議システム 2 は、ネットワーク 20 に收容されたサーバ装置 30 を備える点において、第 1 実施例に係る遠隔会議システム 1 と相違している。
- [0175] サーバ装置 30 は、音声送受信装置 10A 及び 10B を仲介するコンピュータシステムであり、本発明に係る「サーバ装置」の一例である。
- [0176] サーバ装置 30 は、第 1 実施例における処理遅延量決定部 600 を備えており、第 1 及び第 2 計算モデルに基づいた処理遅延量 d 、最大許容処理遅延量 d_{max} 及び各音声送受信装置のバッファ容量に係る各算出処理は、全てこのサーバ装置 30 で実行される構成となっている。
- [0177] 一方、サーバ装置 30 は、決定されたバッファ容量を各音声送受信装置に対し告知する告知部を備えており、告知部からの告知を受けた各音声送受信装置では、バッファ制御部 900 が、この告知されたバッファ容量に従って送受信の各バッファを制御する。
- [0178] このように、第 2 実施例においても、第 1 実施例と同様、快適性を損なわ

ない会話品質が提供される。特に、計算モデルに基づいた各種演算処理やバッファ容量の分配比率決定処理等、比較的高負荷な処理は、音声送受信装置に代わってこのサーバ装置 30 が負担する構成を採るため、音声送受信装置の負担が軽減され、より円滑な会話の運用制御が可能となる。

[0179] 本発明は、上述した実施例に限られるものではなく、請求の範囲及び明細書全体から読み取れる発明の要旨或いは思想に反しない範囲で適宜変更可能であり、そのような変更を伴う音声送受信装置、音声送受信システム及びサーバ装置もまた本発明の技術的範囲に含まれるものである。

産業上の利用可能性

[0180] 本発明は、遠隔地のユーザ同士で会話を成立させる装置或いはシステムに適用可能である。

符号の説明

[0181] 1…遠隔会議システム、10A、10B…音声送受信装置、20…ネットワーク、100…音声入力ユニット、110…音声入力部、120…エンコーダ、130…送信バッファ、140…音声データ送信部、200…音声出力ユニット、210…音声データ受信部、220…受信バッファ、230…デコーダ、240…音声出力部、300…会話特徴量検出部、400…会話特徴量統計処理部、500…記憶部、600…処理遅延量決定部、610…許容処理遅延推定部、620…交渉部、700…RTT測定部、800…処理遅延情報通信部、900…バッファ制御部。

請求の範囲

[請求項1]

各々がネットワークに收容される、複数の拠点に設置された音声送受信装置を含み、該音声送受信装置を介して、前記複数の拠点に夫々存在するユーザ相互間で会話を成立させることが可能な音声送受信システムにおける、前記音声送受信装置であって、

前記複数の拠点のうち自拠点を除く他拠点に設置される前記音声送受信装置との間で、前記ネットワークを介して、少なくとも音声データを含む、前記会話の内容を表す会話データの送受信を行う通信手段と、

前記自拠点到存在する自拠点ユーザ及び前記他拠点到存在する他拠点ユーザの各々について、前記会話における発話のタイミングに関連する所定の会話特徴量を取得する会話特徴量取得手段と、

前記取得された会話特徴量に基づいて、その大小が前記会話の時間遅延量の大小に夫々対応し且つ前記会話の音声品質の高低に夫々対応する前記会話データの処理遅延量の、前記音声送受信システム全体における総量を決定する処理遅延量決定手段と、

前記決定された総量が満たされるように前記処理遅延量を制御する処理遅延量制御手段と

を具備することを特徴とする音声送受信装置。

[請求項2]

前記会話特徴量取得手段は、前記会話特徴量として、前記自拠点到において前記他拠点ユーザの発話出力が終了した時点から前記自拠点ユーザが発話を開始する時点までの時間である自拠点ユーザ反応時間と、前記他拠点点到において前記自拠点ユーザの発話出力が終了した時点から前記他拠点ユーザが発話を開始する時点までの時間である他拠点ユーザ反応時間とのうち少なくとも一方を取得する

ことを特徴とする請求の範囲第1項に記載の音声送受信装置。

[請求項3]

前記処理遅延量決定手段は、前記取得された他拠点ユーザ反応時間の大小に応じて夫々大小に変化する自拠点第1許容処理遅延量及び前

記取得された自拠点ユーザ反応時間の大小に応じて夫々大小に変化する他拠点第1許容処理遅延量のうち最小となる値以下の範囲で前記総量を決定する

ことを特徴とする請求の範囲第2項に記載の音声送受信装置。

[請求項4]

前記通信手段は、前記他拠点に設置される音声送受信装置に対し、前記ネットワークを介して、前記取得された自拠点ユーザ反応時間に対応する自拠点ユーザ反応時間データを送信する

ことを特徴とする請求の範囲第2項に記載の音声送受信装置。

[請求項5]

前記会話特徴量取得手段は、前記会話特徴量として、前記自拠点において前記自拠点ユーザが発話を終了した時点から再び前記自拠点ユーザが発話を開始する時点までの時間である自拠点ユーザ反応待機時間を更に取得する

ことを特徴とする請求の範囲第2項に記載の音声送受信装置。

[請求項6]

前記取得された自拠点ユーザ反応待機時間と前記取得された他拠点ユーザ反応時間との差の大小に応じて夫々大小に変化する自拠点第2許容処理遅延量及び前記他拠点において前記他拠点ユーザが発話を終了した時点から再び前記他拠点ユーザが発話を開始する時点までの時間として取得された他拠点ユーザ反応待機時間と前記取得された自拠点ユーザ反応時間との差の大小に応じて夫々大小に変化する他拠点第2許容処理遅延量のうち最小となる値以下の範囲で前記総量を決定する

ことを特徴とする請求の範囲第5項に記載の音声送受信装置。

[請求項7]

前記処理遅延量決定手段は、前記取得された他拠点ユーザ反応時間の大小に応じて夫々大小に変化する自拠点第1許容処理遅延量及び前記取得された自拠点ユーザ反応時間の大小に応じて夫々大小に変化する他拠点第1許容処理遅延量並びに前記自拠点第2許容処理遅延量及び前記他拠点第2許容処理遅延量のうち最小となる値以下の範囲で前記総量を決定する

ことを特徴とする請求の範囲第 6 項に記載の音声送受信装置。

[請求項8]

前記音声送受信装置は、

前記会話データを前記送受信に相前後して一時的に蓄積するバッファを具備し、

前記処理遅延量制御手段は、前記決定された総量に基づいて、前記バッファに係るバッファ容量を制御する

ことを特徴とする請求の範囲第 1 項に記載の音声送受信装置。

[請求項9]

前記処理遅延量制御手段は、前記決定された総量に基づいて、前記会話データを符号化するにあたっての符号化レートを制御する

ことを特徴とする請求の範囲第 1 項に記載の音声送受信装置。

[請求項10]

前記ネットワークの伝送状態を取得する伝送状態取得手段を更に具備し、

前記処理遅延量決定手段は、前記取得された会話特徴量と前記取得された伝送状態とに基づいて前記総量を決定する

ことを特徴とする請求の範囲第 1 項に記載の音声送受信装置。

[請求項11]

前記伝送状態取得手段は、前記ネットワークの伝送状態として、前記ネットワークの伝送遅延量を取得する

ことを特徴とする請求の範囲第 10 項に記載の音声送受信装置。

[請求項12]

前記取得された会話特徴量を統計処理する統計処理手段を更に具備し、

前記処理遅延量決定手段は、前記統計処理された会話特徴量に基づいて前記総量を決定する

ことを特徴とする請求の範囲第 1 項に記載の音声送受信装置。

[請求項13]

前記取得された会話特徴量を記憶する記憶手段を更に具備する

ことを特徴とする請求の範囲第 1 項に記載の音声送受信装置。

[請求項14]

各々がネットワークに收容される、複数の拠点に設置された音声送受信装置を含み、該音声送受信装置を介して、前記複数の拠点に夫々存在するユーザ相互間で会話を成立させることが可能な音声送受信シ

ステムであって、

前記音声送受信装置は、

前記複数の拠点のうち自拠点を除く他拠点に設置される前記音声送受信装置との間で、前記ネットワークを介して、少なくとも音声データを含む、前記会話の内容を表す会話データの送受信を行う通信手段と、

前記自拠点到存在する自拠点ユーザ及び前記他拠点到存在する他拠点ユーザの各々について、前記会話における発話のタイミングに関連する所定の会話特徴量を取得する会話特徴量取得手段と、

前記取得された会話特徴量に基づいて、その大小が前記会話の時間遅延量の大小に夫々対応し且つ前記会話の音声品質の高低に夫々対応する前記会話データの処理遅延量の、前記音声送受信システム全体における総量を決定する処理遅延量決定手段と、

前記決定された総量が満たされるように前記処理遅延量を制御する処理遅延量制御手段と

を具備することを特徴とする音声送受信システム。

[請求項15]

各々がネットワークに收容される、サーバ装置と、複数の拠点到設置された、該複数の拠点のうち自拠点を除く他拠点到設置される音声送受信装置との間で、前記ネットワークを介して、少なくとも音声データを含む、前記会話の内容を表す会話データの送受信を行う通信手段を具備する音声送受信装置とを含み、該音声送受信装置を介して、前記複数の拠点到夫々存在するユーザ相互間で会話を成立させることが可能な音声送受信システムにおける、前記サーバ装置であって、

前記複数の音声送受信装置から、前記ネットワークを介して、前記自拠点到存在する自拠点ユーザ及び前記他拠点到存在する他拠点ユーザの各々についての前記会話における発話のタイミングに関連する所定の会話特徴量を取得する会話特徴量取得手段と、

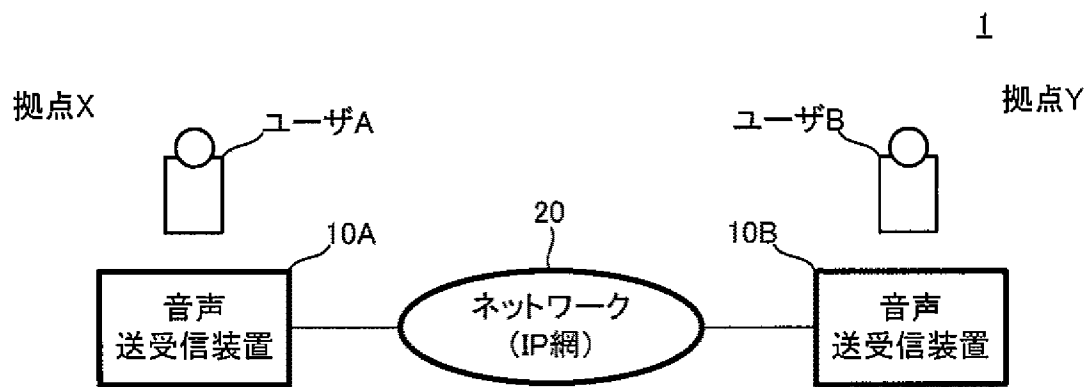
前記取得された会話特徴量に基づいて、その大小が前記会話の時間

遅延量の大小に夫々対応し且つ前記会話の音声品質の高低に夫々対応する前記会話データの処理遅延量の、前記音声送受信システム全体における総量を決定する処理遅延量決定手段と、

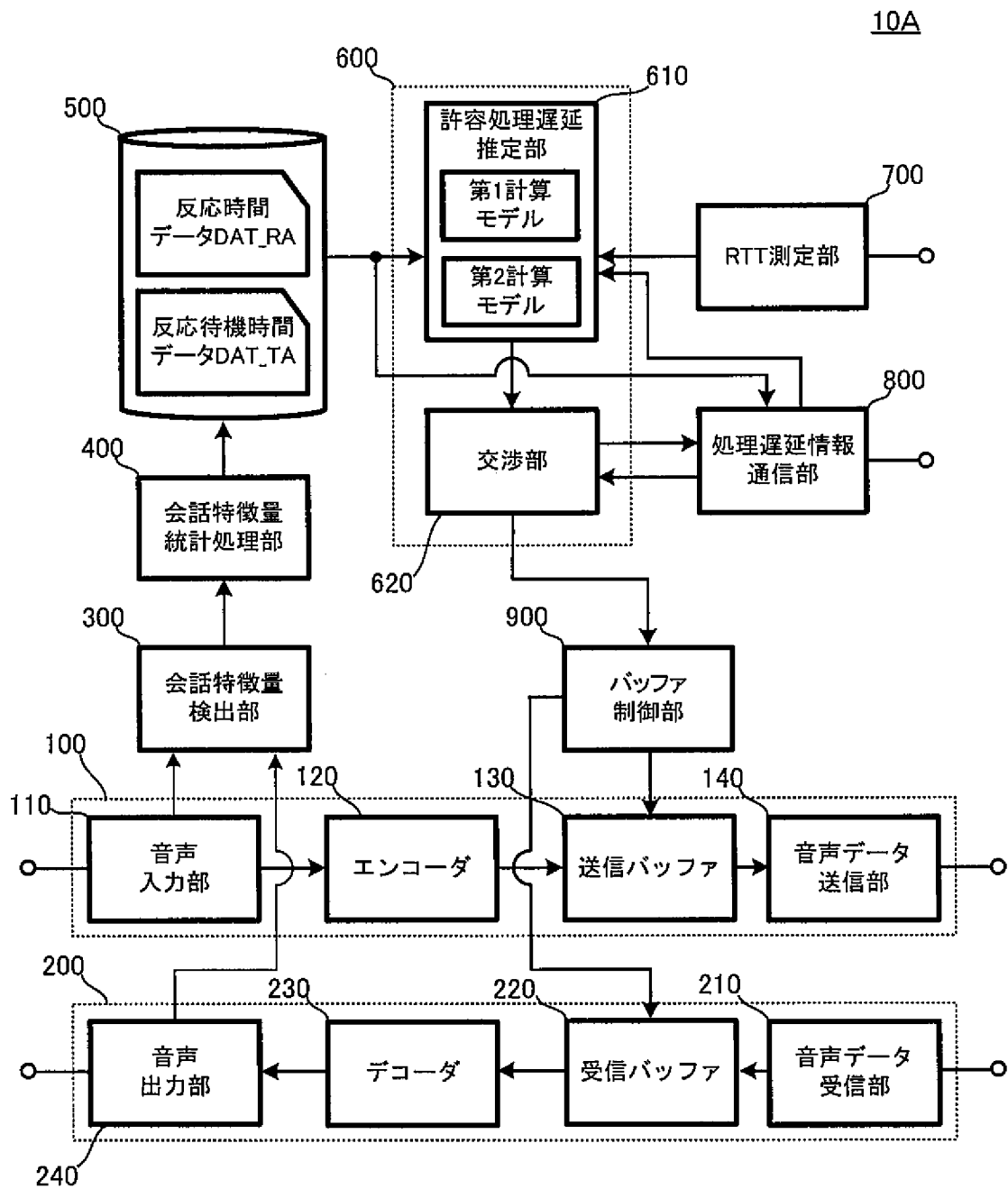
前記複数の音声送受信装置に対し前記ネットワークを介して前記決定された総量を告知する告知手段と

を具備することを特徴とするサーバ装置。

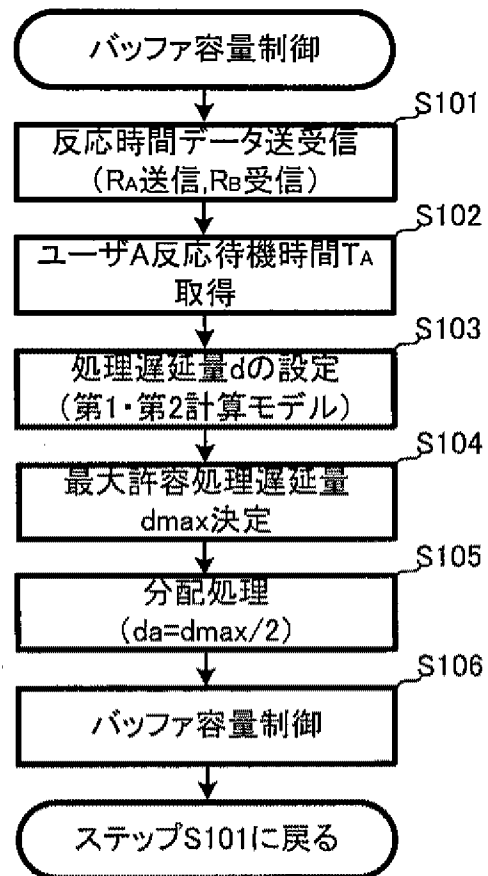
[図1]



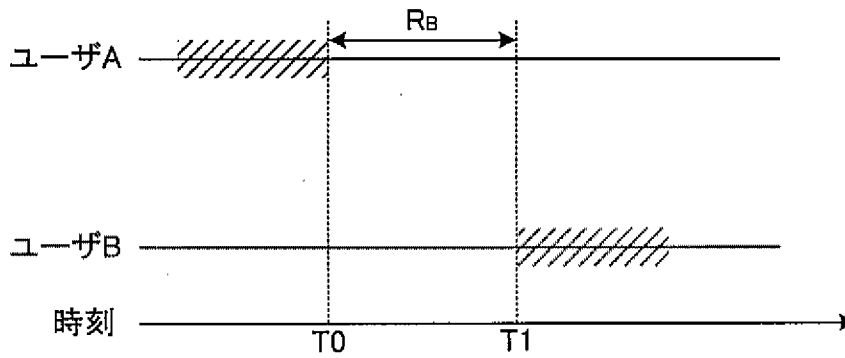
[図2]



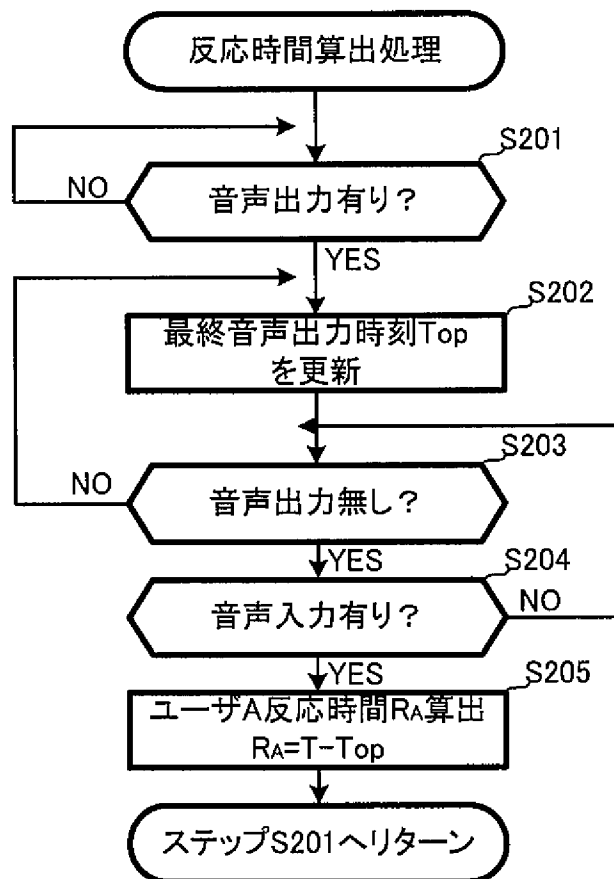
[図3]



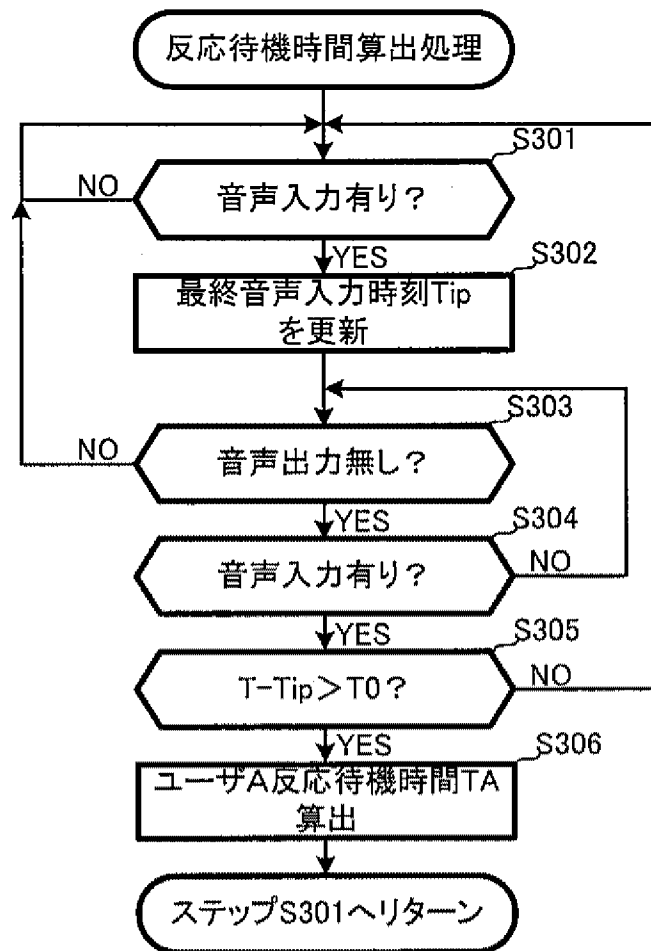
[図4]



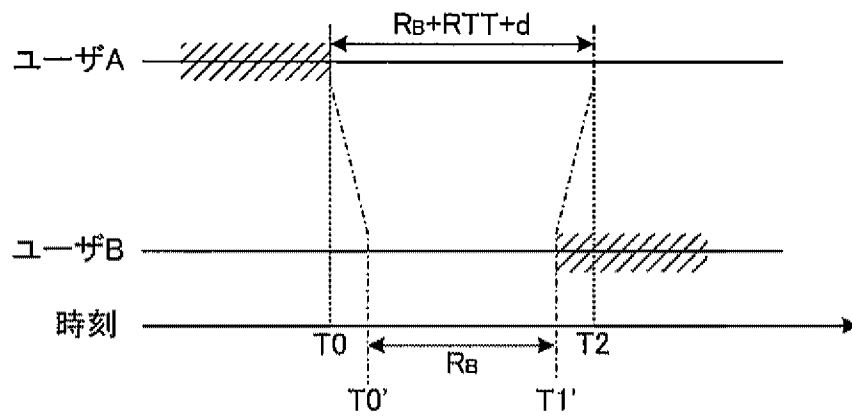
[図5]



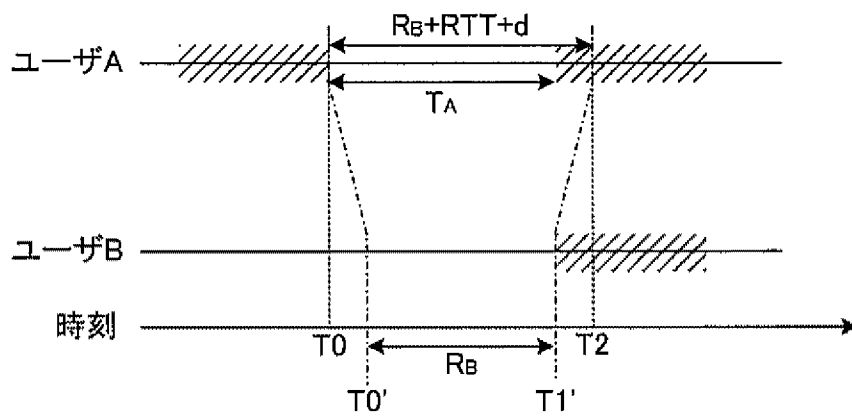
[図6]



[図7]

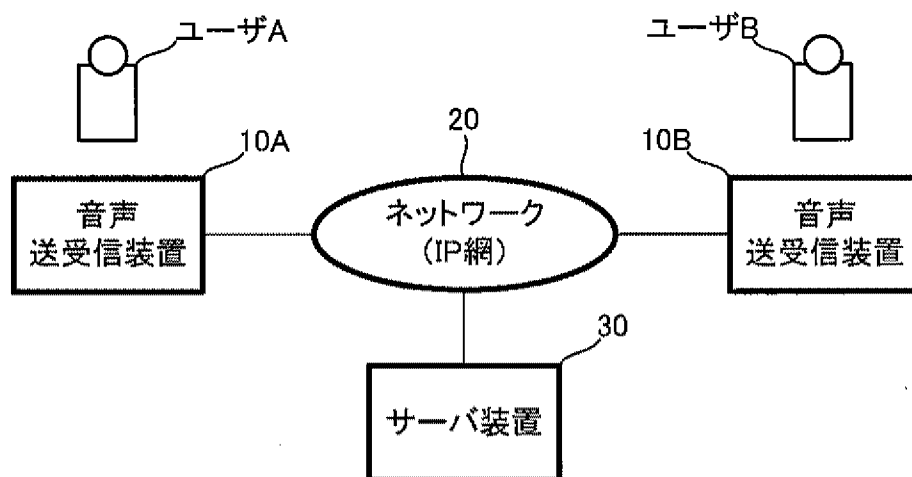


[図8]



[図9]

2



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/062558

A. CLASSIFICATION OF SUBJECT MATTER

H04L12/56(2006.01) i, H04M11/06(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L12/56, H04M11/06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2010
Kokai Jitsuyo Shinan Koho	1971-2010	Toroku Jitsuyo Shinan Koho	1994-2010

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2002-164921 A (Oki Electric Industry Co., Ltd.), 07 June 2002 (07.06.2002), paragraphs [0121] to [0157] & US 2002/0064158 A1	1-15
Y	JP 2005-303531 A (Mitsubishi Electric Corp.), 27 October 2005 (27.10.2005), claim 1 (Family: none)	1-15
A	JP 10-200580 A (Matsushita Electric Industrial Co., Ltd.), 31 July 1998 (31.07.1998), claim 1 (Family: none)	1-15

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
24 September, 2010 (24.09.10)

Date of mailing of the international search report
05 October, 2010 (05.10.10)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. H04L12/56(2006.01)i, H04M11/06(2006.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int.Cl. H04L12/56, H04M11/06

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2010年
日本国実用新案登録公報	1996-2010年
日本国登録実用新案公報	1994-2010年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	JP 2002-164921 A (沖電気工業株式会社) 2002.06.07, 段落【0121】～【0157】 & US 2002/0064158 A1	1-15
Y	JP 2005-303531 A (三菱電機株式会社) 2005.10.27, 【請求項1】 (ファミリーなし)	1-15
A	JP 10-200580 A (松下電器産業株式会社) 1998.07.31, 【請求項1】 (ファミリーなし)	1-15

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

24.09.2010

国際調査報告の発送日

05.10.2010

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

齋藤 浩兵

電話番号 03-3581-1101 内線 3596

5 X

3794