

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2006-238445
(P2006-238445A)

(43) 公開日 平成18年9月7日(2006.9.7)

(51) Int. Cl.	F I	テーマコード (参考)
HO4L 12/56 (2006.01)	HO4L 12/56 230A	5K030
HO4M 3/00 (2006.01)	HO4M 3/00 B	5K201

審査請求 未請求 請求項の数 10 O L (全 15 頁)

(21) 出願番号 特願2006-44971 (P2006-44971)
 (22) 出願日 平成18年2月22日 (2006.2.22)
 (31) 優先権主張番号 11/062966
 (32) 優先日 平成17年2月22日 (2005.2.22)
 (33) 優先権主張国 米国 (US)

(71) 出願人 596092698
 ルーセント テクノロジーズ インコーポ
 レーテッド
 アメリカ合衆国, 07974-0636
 ニュージャージー, マレイ ヒル, マウン
 テン アヴェニュー 600
 (74) 代理人 100064447
 弁理士 岡部 正夫
 (74) 代理人 100085176
 弁理士 加藤 伸晃
 (74) 代理人 100094112
 弁理士 岡部 譲
 (74) 代理人 100096943
 弁理士 臼井 伸一

最終頁に続く

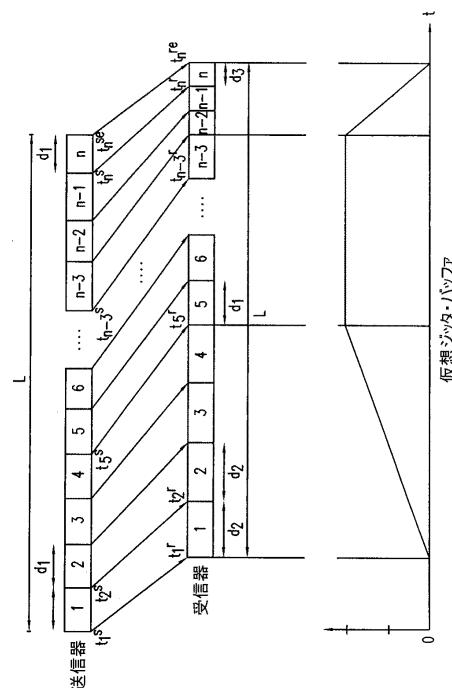
(54) 【発明の名称】 仮想ジッタ・バッファおよび時間スケール修正を使用した、Voice over IP通信ネットワークにおいてネットワーク・ジッタを処理する方法および装置

(57) 【要約】

【課題】 仮想ジッタ・バッファおよび時間スケール修正を使用した、Voice over IP通信ネットワークにおいてネットワーク・ジッタを処理する方法および装置を提供すること。

【解決手段】 音声パケットのシーケンスがトーク・スパートを含むことができる、音声パケットのシーケンスが受信され、パケットは、ゼロの初期再生レーテンシを有するジッタ・バッファに記憶される。その後、シーケンス内の複数の初期音声パケットが、有利には、時間伸張され、再生出力され、それによって、ジッタ・バッファの深さ（処理されるネットワーク・ジッタ量）を増加させる。その後、ジッタ・バッファの深さをゼロに戻すために、複数の後続の音声パケットが、時間圧縮され、再生出力される。一実施形態では、音声パケットのシーケンスは、トーク・スパートを含み、別の実施形態では、音声パケットのシーケンスは、携帯のハンドオフ中に受信される音声パケットを含む。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

通信ネットワーク内の受信器においてネットワーク・ジッタを処理する方法であって、前記受信器は、前記通信ネットワークを通じて受信した、音声パケットの対応するシーケンス内に含まれる音声データのフレームのシーケンスを記憶するジッタ・バッファを有しており、方法は、

前記対応する音声パケットを受信すると、前記ジッタ・バッファ内に前記音声データのフレームのそれぞれを記憶する工程と、

前記音声データのフレームのシーケンスの第 1 サブシーケンスを識別する工程と、

前記音声データのフレームの前記第 1 サブシーケンスにおいて、前記音声データの記憶されたフレームのそれぞれに関して時間伸張を実施し、前記音声データの時間伸張されたフレームのそれぞれを、連続して再生出力する工程と、

前記音声データのフレームのシーケンスの第 2 サブシーケンスを識別する工程と、

前記音声データのフレームの前記第 2 サブシーケンスにおいて、前記音声データの記憶されたフレームのそれぞれに関して時間圧縮を実施し、前記音声データの時間圧縮されたフレームのそれぞれを連続して再生出力する工程とを含む方法。

10

【請求項 2】

前記音声データのフレームのシーケンスはトーク・スパートを含み、前記音声データのフレームのシーケンスの前記第 1 サブシーケンスは、前記トーク・スパート内に含まれる音声データの 1 つまたは複数の初期フレームを含み、前記音声データのフレームのシーケンスの前記第 2 サブシーケンスは、前記トーク・スパート内に含まれる音声データの 1 つまたは複数の終了フレームを含む請求項 1 に記載の方法。

20

【請求項 3】

前記音声データのフレームのシーケンス内の前記音声データのフレームの第 1 サブシーケンスに続いて、かつ、前記音声データのフレームのシーケンス内の前記音声データのフレームの第 2 サブシーケンスの前に、前記音声データの記憶されたフレームの 1 つまたは複数、標準速度で再生出力する工程をさらに含む請求項 2 に記載の方法。

【請求項 4】

前記音声データのフレームのシーケンスの前記第 1 サブシーケンスは、一定の所定数のフレームを含み、前記フレームの前記時間伸張は、一定の時間スケール比によって前記フ

30

【請求項 5】

前記一定の所定数のフレームは 4 に等しく、前記フレームが時間伸張される前記一定の時間スケール比は 50% に等しい請求項 4 に記載の方法。

【請求項 6】

前記トーク・スパート内の前記音声データのフレームのシーケンスの最後のフレームに対応する音声パケットの受信を検出する工程をさらに含み、前記音声データのフレームのシーケンスの前記第 2 サブシーケンスは、

(i) 前記ジッタ・バッファに記憶され、前記トーク・スパート内の前記音声データのフレームのシーケンスの前記最後のフレームに対応する前記音声パケットの前記受信の前記検出時点まで、まだ再生出力されていないフレームと、

40

(i i) 前記トーク・スパート内の前記音声データのフレームのシーケンスの前記最後のフレームとを含む請求項 3 に記載の方法。

【請求項 7】

前記トーク・スパート内の前記音声データのフレームのシーケンスの前記最後のフレームに対応する前記音声パケットの前記受信は、半分のレートのフレームが直後に続く、8分の 1 レート・フレームを含む前記対応する音声パケットに基づいて検出される請求項 6 に記載の方法。

【請求項 8】

8分の 1 レート・フレームが直後に続く、フル・レート・フレームを含む前記対応する

50

音声パケットに基づいて、前記トーク・スパート内の前記音声データのフレームのシーケンスの最初のフレームに対応する音声パケットの受信を検出する工程をさらに含む請求項 3 に記載の方法。

【請求項 9】

前記通信ネットワークは無線通信ネットワークを備え、前記受信器はモバイル局を備える請求項 1 に記載の方法。

【請求項 10】

前記音声データのフレームのシーケンスの前記第 1 サブシーケンスは、第 1 基地局から前記モバイル局によって受信される音声パケットに対応し、前記音声データのフレームのシーケンスの前記第 2 サブシーケンスは、第 2 基地局から前記モバイル局によって受信される音声パケットに対応し、前記モバイル局は、前記第 1 基地局から前記第 2 基地局へのハンドオフを受ける請求項 9 に記載の方法。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般に、VoIP (Voice over Internet Protocol) 通信サービスを提供するパケット・ベースの通信ネットワークの分野に関し、特に、付加的な再生レテンシを導入することなく、ネットワーク・ジッタを処理する方法および装置に関する。

【背景技術】

20

【0002】

音声パケットが、VoIP (Voice over Internet Protocol) を提供する通信ネットワークを通して伝送される時に、連続するパケットの受信間の間隔の規則性が必然的に乱される。この現象は、「ネットワーク・ジッタ」として知られている。こうしたジッタは、パケットがあまりにも遅く到着する時はいつでも、パケット喪失をもたらす可能性があり、それによって、その再生時間がなくなり、これらのパケット喪失は、最終的には音声品質を低下させる。典型的な最先端の VoIP システムでは、受信側エンド上のパケット・バッファは、ジッタを補償するために、一定量の初期再生遅延を付加することによって、この問題を解消するのに使用される。この「ジッタ・バッファ」は、その後、一定遅延後に開始する、音声パケットの安定したストリームを再生メカニズム内に供給する。そのため、一定遅延量までの任意のジッタ量が、音声品質の損失無しで、対処されるであろう。伝送におけるその遅延が、この一定遅延を超えるパケットのみが、「喪失される」ことになる。(こうして「喪失された」パケットについて、ほとんどのシステムは、結果として生じる音声品質の低下を制限しようとして、従来のパケット喪失隠蔽技法を採用する。それでも、過度のパケット喪失は、かなりの音声品質低下を必然的にもたらすであろう。)

30

【0003】

したがって、ジッタ・バッファの初期レテンシが任意に長い場合、任意に大きな量のネットワーク・ジッタが、理論的には吸収されるであろう。しかし、実世界の音声通信システムでは、リアルタイムの制約により、長い遅延は実用的ではない。これらの実世界の制約により、ほとんどの音声パケット通信システムの典型的なジッタ・バッファの初期再生レテンシは、20~60ミリ秒であり、それによって、20~60ミリ秒の付加的なエンド-エンド遅延をもたらす。ほとんどのジッタ・バッファ設計では、この初期再生レテンシは、バッファによって与えられるジッタ保護と同じである。そのため、増加したジッタに対処するために、再生レテンシは、それに応じて、増加しなければならない。VoIP、特に、無線 VoIP において、この付加的な遅延は、ネットワーク容量、音声品質などの点で、重要な役割を果たすことができる。

40

【0004】

ランド-モバイル接続では、ジッタ・バッファが、たとえ 20 ミリ秒遅延のみを挿入しても、パケット喪失がほんのわずかである傾向がある。しかし、モバイル-ランドおよび

50

モバイル-モバイル接続では、ジッタは、実質的により顕著であり、パケット喪失を許容可能なレベルにするには、60ミリ秒以上の遅延が必要とされることが多い。ネットワークの詳細が、観測される正確なジッタ量に影響を与えるが、少なくとも数パケット長のジッタは、適度な音声品質を送出するために対処される必要があるという一般的な傾向が当てはまる。(音声パケットは、通常、20ミリ秒のスピーチを含む。)さらに、所与のモバイル基地局上のユーザの数が増加すると、ジッタが大幅に増加するため、受信器において増加するジッタ量に対処する能力は、有利には、各基地局上で、より多くの数のユーザが適度の音声品質によってサポートされることをもたらさずであろう。

【0005】

トーカが、音声通信システムの電話を通じて話す時、メッセージは、基礎をなすネットワークを通して他の側に送出される。ITU-T G.114という、国際電気通信連合(International Telecommunication Union)によって公表され、当業者に熟知されている、よく知られている規格は、この規格に準拠するネットワーク内の2人の人間ユーザの間での通常のテレフォニー会話に許可される最大エンド-エンド遅延を指定する。特に、総エンド-エンド遅延は、150ミリ秒より小さいことが推奨される。しかし、たとえば、CDMA 2000 EVDO rev. Aという、当業者に熟知されている無線通信規格などの無線VoIPシステムでは、この遅延要件を一貫して達成することは、非常に難しいことが多い。その理由は、従来のジッタ・バッファを使用すると、高いパケット・ジッタが、大きな一定遅延(高い再生レテンシから生じる)に直接変換されるためである。

10

20

【0006】

採用されてきた一手法は、「適応」ジッタ・バッファを使用することであり、一定遅延量は、たとえば、トーク・スパートごとにトーク・スパートに関してなどで、調整されてもよい。(「トーク・スパート」は、無音が先行し、かつ、無音が後に続くトーカのスピーチのセグメントである。)こうした手法は、ネットワークの振る舞い(たとえば、前の一定遅延量がジッタ・バッファによって付加されるとした場合のパケット喪失量)を調べ、その後、それに応じて、次のトーク・スパートについてジッタ・バッファを調整する。しかし、この手法は、たとえ、非常に適応的にネットワーク条件に基づいたとしても、ジッタ保護量に等しい再生レテンシ量をやはり導入する。

30

【発明の開示】

【発明が解決しようとする課題】

【0007】

したがって、高い再生レテンシを導入することなく、高いネットワーク・ジッタを処理する方法および装置を提供することが、非常に望ましいであろう。

【課題を解決するための手段】

【0008】

本発明の原理によれば、仮想ジッタ・バッファおよび時間スケール修正を使用した、通信ネットワークにおいてネットワーク・ジッタを処理する方法および装置が提供される。本発明の1つの説明的な実施形態によれば、音声パケットのシーケンスが好ましくはトーク・スパートを含むことができる、音声パケットのシーケンスが受信され、パケットは、ゼロの初期再生レテンシを有するジッタ・バッファに記憶され、それによって、効果的にジッタ・バッファをスキップし、リスナに対して、遅延無しで即座にパケットを再生出力する。所定の数の初期音声パケット(たとえば、シーケンスの最初の数個の音声パケット)は、有利には、再生出力される時に時間伸張され、それによって、所定の最大有効レテンシに達するまで、ジッタ・バッファの有効レテンシ(したがって、処理されるネットワーク・ジッタ)を増加させる。その後、トーク・スパートの終了を指示するパケットの受信が検出されるまで、後続の音声パケットが、その標準的な長さ(すなわち、標準速度)で再生出力され、その時点で、残りの(未再生出力の)音声パケットは、有利には、有効ジッタ・バッファ・レテンシをゼロに戻すために、再生出力される時に時間圧縮される。

40

50

【0009】

こうして、実質的にゼロの遅延ジッタ・バッファが、有利には、達成され、一方、多くの音声パケットは、それでも、所定の量だけネットワーク・ジッタから保護される。すなわち、本発明のこの説明的な実施形態によれば、仮想ジッタ・バッファは、トーク・スパークの「中央」でネットワーク・ジッタ保護を提供し、それによって、大多数の音声パケットをネットワーク・ジッタから保護し、なおかつ、仮想ジッタ・バッファ長は、各トーク・スパークの開始および終了においてゼロになり、それによって、遅延の付加が知覚されないことになる。

【0010】

本発明の別の説明的な実施形態によれば、仮想ジッタ・バッファは、有利には、ネットワーク・ジッタが最も問題となる可能性がある、携帯の「ハンドオフ」が起こる時に、モバイル（たとえば、無線）通信ネットワークにおいて付加されたネットワーク・ジッタ保護を提供し、なおかつ、仮想ジッタ・バッファ長は、ネットワーク・ジッタがあまり問題とならない可能性がある、他の時には増加しない。特に、モバイル局が、サービス基地局を、1つの基地局から別の基地局へ切り換える期間は、通常、サービス・ギャップを作成する。本発明のこの説明的な実施形態によれば、このサービス・ギャップは、ハンドオフを予想しての時間伸張およびハンドオフが完了した後の時間圧縮の有利な使用によって埋められる。

10

【発明を実施するための最良の形態】

【0011】

20

[本発明の第1の説明的な実施形態]

本発明の1つの説明的な実施形態によれば、トーカー間の各会話ターンにおいて、トーク・スパークを積極的に管理する方法が採用される。説明的なトーク・スパーク管理方法は、有利には、各トーク・スパーク内でパケットにジッタ保護を提供しながら、ゼロ初期バッファリング遅延を可能にし、それによって、パケット喪失率の減少、相応して、音声品質の向上をもたらす。図1は、従来技術のジッタ・バッファ技法が採用される時の、2人のスピーカ間の会話ターン、および、結果として得られるスピーカ間の会話遅延のサンプルタイミング図を示す。スピーカは、それぞれ、「東」および「西」として識別される。特に、図は、東による第1の会話ターンと、それに続く、西による第1の会話ターンと、それに続く、東による第2の会話ターンと、その後の、西による第2の会話ターンを示す。 (「会話」ターンは、会話に対して、1人の相手のみがある期間である。)

30

【0012】

従来技術の技法によれば、東による第1の会話ターン中に起こるネットワークの振る舞い（たとえば、パケット喪失）の結果として、西の受信器に関連する適応ジッタ・バッファは、東による第2の会話ターンで使用するために、ジッタ・バッファ遅延を付加することが必要であると判断することが、図1で仮定される。そのため、図を見てわかるように、付加されたこのジッタ・バッファ遅延は、東の知覚される遅延の増加をもたらす。

【0013】

各会話ターンは、1つまたは複数のトーク・スパークからなる。一方向伝送遅延（たとえば、図1において、東から西、または、西から東）は、送信器におけるパケット化および符号化遅延、ならびに受信器におけるネットワーク伝送遅延、ジッタ・バッファ遅延、および復号化遅延を含む。本発明の原理によれば、ジッタ・バッファ遅延は、有利には、減少するか、または、なくなる。ジッタ・バッファ遅延を減少させることによって、各トーカーの総合の知覚される遅延もまた、有利には、減少する可能性がある。特に、各会話ターン内での積極的なトーク・スパークの管理によって、ジッタ・バッファ遅延を大幅に減少させることが可能になる。

40

【0014】

より詳細には、本発明の1つの説明的な実施形態によれば、積極的なトーク・スパークの管理は、復号化されるスピーチ・サンプルに対して時間スケール修正技法を適用する。図2は、本発明の1つの説明的な実施形態による、仮想ジッタ・バッファの使用によって

50

単一トーク・スパートの積極的なトーク・スパート管理を実施する時間スケール修正技法の動作を示す。図は、時間スケール修正の結果ならびに使用される仮想ジッタ・バッファの対応する有効再生レーテンシの両方を示す。

【0015】

特に、図は、送信器から受信器に伝送される、数 n の音声パケットそれぞれが d_1 に等しい長さ（すなわち、継続時間）を有するに符号化された、単一トーク・スパートの動作を示す。図に示すように、送信器は、第1音声パケット（すなわち、パケット1）を時間 t_1^s に送信すること、および、そのパケットが、受信器に時間 t_1^r に到着することが仮定される。本発明の説明的な実施形態の動作によれば、伝送遅延 $t_1^r - t_1^s$ を減少させるために、トーク・スパートの第1音声パケット（すなわち、パケット1）は、到着するとすぐに復号され、再生バッファに送出される。（ほとんどの従来技術のシステムでは、パケットは、ネットワーク・ジッタに対するある程度の保護を与えるために、ジッタ・バッファに置かれることになることに留意されたい。）本発明の説明的な実施形態の動作によれば、第1パケットは、即座に再生出力されるため、リスナは、ジッタ・バッファの結果としていずれの付加的な遅延にも気づかない。

10

【0016】

しかし、本発明の説明的な実施形態の動作によれば、第1パケットからのスピーチ・サンプルの長さは、有利には、 d_1 のその元の長さ（すなわち、継続時間）から d_2 (d_2 は d_1 より長い)へ、説明的に、元の長さ d_1 の予め決めたパーセンテージだけ時間伸張される。図に示す例では、 d_2 は、説明的に、 d_1 より50%長い。すなわち、 d_1 が20ミリ秒の長さ（音声パケットにとって典型的な長さ）であった場合、 d_2 は、有利には、30ミリ秒の長さであることになる。そのため、この説明的な例では、それぞれの時間伸張されたパケットは、後続のパケットに対して付加的な10ミリ秒に相当するジッタ保護を提供する。

20

【0017】

第2パケットが到着すると（付加的なジッタが無い状態で到着すると仮定すると）、第2パケットは、復号されることができない前に、キュー内で10ミリ秒待たなければならない。次のパケットは（同じ仮定の下で）、20ミリ秒待つなどである。付加的なサンプルは、「遅い」パケットが関係する限り、送信器と受信器の「クロック」間に時間ラグを生じる。図2に示す説明的な例は、最初の4つのパケットが、50%の時間スケール伸張によって再生出力される場合を示す。それによって、蓄積された総合の40ミリ秒の時間ラグは、ネットワーク・ジッタに対する付加的な保護を提供する。そのため、本発明者等は、本明細書で、この時間スケール伸張の有効な結果を、「仮想ジッタ・バッファ」の作成と呼ぶ。時間スケール伸張はまた、有利には、到来する音声パケット・ストリームの予見能力を提供することに留意されたい。

30

【0018】

本発明の説明的な実施形態の動作によれば、この時間スケール伸張動作は、有利には、時間ラグが予め決めた最大値に達するまで実施される。図2に示す説明的な例では、この予め決めた最大値は40ミリ秒である。仮想ジッタ・バッファのサイズがこの最大値に達すると、後続のパケットは、有利には、トーク・スパートの最後のパケットが受信器によって検出されるまで、すなわち、この最後のパケットが、仮想ジッタ・バッファに「入る」まで、元の時間スケールで再生される。図2に示す説明的な例では、仮想ジッタ・バッファは、有利には、2つのパケットに相当する予見能力を提供する。すなわち、受信器がパケット $n-2$ を復号しようとしている時、受信器は、既に、パケット n （すなわち、最後のパケット）を受信してしまっている。

40

【0019】

本発明の説明的な実施形態によれば、受信されたパケット（この説明的な例では、パケット n ）が、現在のトーク・スパートが終了することを指示すると（こうした判断を行う説明的な技法のために以下を参照されたい）、仮想ジッタ・バッファ内に残っているパケット（すなわち、受信されたが、まだ再生出力されていないパケット）の全て、ならびに

50

、この最後に受信されたパケット（パケット n ）は、復号され、再生出力されるが、有利には、単一パケットの元の長さ d_1 内に全てが再生出力される。こうして、トーク・スパートの終了は、付加される遅延がゼロに保たれ、したがって、ジッタ・バッファは、会話遅延を増やさない。

【0020】

本発明の説明的な実施形態によれば、上述した短縮された再生間隔は、有利には、時間スケール圧縮の使用によって達成される。説明的に、各圧縮されたパケットの圧縮長 d_3 は、有利には、 $d_1 / (\text{仮想ジッタ・バッファ・サイズ} + 1)$ に等しくなるように計算されてもよい。この説明的な例では、 d_3 は6.6667ミリ秒である。こうして、受信器のクロックは、有利には、送信器のクロックに「追い付く」ことになる、または、等価的に、仮想ジッタ・バッファは、ゼロ・サイズに戻るようになる。その結果、全体のトーク・スパート長（図では L として示す）は、有利には、送信器と受信器の両方について同じである。この条件は、有利には、会話遅延を最小にすることをもたらすことに留意されたい。同様に、図に説明的に示した、最初の4つのパケットのそれぞれについての10ミリ秒の付加的なラグは、いずれの気づかれる音声品質（または、韻律素）の低下ももたらさず、一方、上述した利点の全てをやはり提供することに留意されたい。

10

【0021】

図3は、本発明の1つの説明的な実施形態による、仮想ジッタ・バッファの使用によって、会話ターンの積極的なトーク・スパートの管理を実施する時間スケール修正技法の動作を示す。図を見てわかるように、説明的な会話ターンは3つのトーク・スパートからなる。図は、送信器および受信器についてのタイミング・シーケンスならびに使用される仮想ジッタ・バッファの対応する有効深さの両方を示す。（都合のために、送信器および受信器のタイミング・シーケンスは、伝送遅延を無視して、図に示される。）

20

【0022】

多くの可変レート・ボコーダ（音声コーダ）は、スピーチの有声の/無声の、および、無音の部分について異なるパケット・タイプを使用する。たとえば、よく知られているEVR Cボコーダという、当業者に熟知されている従来のCDMA無線通信システムで使用される標準的な音声コーデック（コーダ/デコーダ）は、可変符号化レート・アクティブ・スピーチ・パケット用のフルレート符号化およびハーフレート符号化、ならびに、無音間隔パケット用の1/8レート符号化を使用する。図3に示す説明的な例の議論において、本発明者等は、こうしたEVR Cボコーダの使用を仮定するであろう。

30

【0023】

本発明の説明的な実施形態によれば、会話ターンの開始および終了は、有利には、先行する1/8レート（すなわち、無音）パケットの数をカウントすることによって検出されてもよい。1/8レート・パケットが、一部のボコーダ実施態様でそうであるように、抑圧される（すなわち、ブランキングされる）場合、内部クロック、または、当業者に熟知されている、リアルタイム・プロトコル（RTP）などの通信プロトコルは、代わりに、送出されたと思われるこうしたパケットの数を効果的にカウントするのに使用されることができる。

【0024】

そのため、具体的に、本発明の説明的な実施形態によれば、フルレートまたはハーフレートのパケットが、その長さが、ある所定の閾値より大きい、一連の1/8レート・パケットに続く場合、この事実は、有利には、新しい会話ターン、したがって、新しいトーク・スパートの開始を指示するのに使用されてもよい。そのため、最初のトーク・スパートの開始部分は、先に述べ、図2に示すように、有利には、時間伸張され、トーク・スパートの終了部分は、有利には、圧縮される。

40

【0025】

しかし、本発明の1つの説明的な実施形態によれば、任意の後続のトーク・スパート（すなわち、所与の会話ターンの最初のトーク・スパート以外の任意のトーク・スパート）の開始は、必ずしも時間伸張されない。代わりに、それは、付加的な無音間隔の使用によ

50

って、最大仮想ジッタ・バッファまでの時間量だけ単に遅延されてもよい。これは、有利には、これらの後続のトーク・スパートの開始部分をネットワーク・ジッタから最大限保護するために行われ、そうでなければ、それらはジッタ保護をほとんど受けないことになる（会話ターンの最初のトーク・スパートの最初のいくつかのパケットの場合がそうである）。

【0026】

さらに、本発明のこの説明的な実施形態によれば、受信器は、一般に、所与のトーク・スパートの終了が、実際に、会話ターンの終了であるかどうかを知らないため、送信器と再同期化するために、各トーク・スパートの終了部分は、有利には、圧縮される（図2に関連して先に述べたように）。しかし、無音間隔が比較的短く、かつ、したがって、所与のトーク・スパートの終了が、会話ターンを終了させないと仮定することができる場合、後続のトーク・スパートは、有利には、知覚される遅延に影響を及ぼすことなく、遅延されてもよい。本発明者等は、本明細書でこれを、「プッシング無音」と呼ぶ。

10

【0027】

無音間隔が非常に短い場合、無音は、人が、たとえば、「フットボール」という語を普通に言うと思われる言い方などの、単一の語内にあることが、よくあることに留意されたい。この場合、後続のスピーチは、それが、まるで会話ターンの開始であるかのように、伸張される必要がある場合がある。これは、気づくことができる無音ギャップが、語の中央に間違ってプッシュされることを防止する（たとえば、「フット。ボール。」のように聞こえるのではなく、「フットボール」ように聞こえるスピーチをもたらすために）。

20

【0028】

本発明の説明的な実施形態に従って使用される、上述した技法は、有利には、たとえ、全てのトーク・スパートが会話ターンを指示するわけではないにしても、各会話ターンの開始と終了が、ゼロ・ジッタ・バッファ遅延を有することを確実にするであろう。仮想ジッタ・バッファ深さを示す図3において、アクティブ・スピーチ・パケットのほとんど（唯一の例外は、最初のトーク・スパートの開始部分のパケットである）は、最大仮想ジッタ・バッファ深さまで、ジッタ保護されることに留意されたい。特に、図で説明的に示すように、少数の例外のパケット以外の全てのパケットについてのジッタ保護は、いずれの付加的な知覚される遅延も導入することなく、40ミリ秒の再生遅延を有する一定のジッタ・バッファによって与えられる保護に等しい。

30

【0029】

上述した本発明の説明的な実施形態によれば、積極的なトーク・スパートの管理は、有利には、時間スケール修正、すなわち、時間スケール伸張および圧縮の使用によって使用可能にされる。音声パケットの時間スケール伸張は、有利には、パケットに含まれるスピーチのピッチ期間を繰り返すことによって達成されてもよく、時間スケール圧縮は、有利には、パケットに含まれるスピーチのピッチ期間を落とすことによって達成されてもよい。こうした動作は、元のスピーチのピッチを変えずに、パケットからの復号されたスピーチ・サンプルの長さを修正する。

【0030】

先に指摘したように、多くの可変レート・ボコーダは、スピーチの有声の/無声の、および、無音の部分について、異なるパケット・タイプを使用する。本発明の説明的な実施形態によれば、トーク・スパートおよび会話ターンを検出する方法は、ほとんどの可変レート・ボコーダが提供する符号化レート情報を使用する。先に指摘したように、たとえば、E V R C ボコーダは、可変レート、すなわち、アクティブ・スピーチについて、フルレートおよびハーフレートのパケット、および、無音間隔について1/8レート・パケットを使用する。

40

【0031】

図4は、本発明の説明的な実施形態による、トーク・スパート検出を実施する方法のフロー図を示す。図に示す方法は、上述したE V R C ボコーダのレート符号化特性を仮定する。

50

【0032】

特に、図4に示す本発明の説明的な実施形態によれば、新しいトーク・スパートの開始は、有利には、 $nThr1$ (所定の閾値) 以上の無音パケット (たとえば、EVRコードブックの1/8レート・パケット) が存在したかどうかを仮定される。スピーチ・パケット (たとえば、EVRコードブックのフルレートまたはハーフレートのパケット) が受信されると、それは、遅延無しで再生され、仮想ジッタ保護を構築するために伸張される。 $nThr0$ ($nThr0 < nThr1$ である、別の所定の閾値) より少ない無音パケットが受信された場合、無音は、有利には、トーク間無音として解釈され、同じ伸張戦略が、有利には、使用される。(説明的に、 $nThr0$ は3に等しくてもよく、 $nThr1$ は15に等しくてもよい。) これは、無音を語の中央にプッシュすることを回避する。(パケット内の) 無音間隔の長さが、2つの閾値 $nThr0$ と $nThr1$ の間にある場合、無音は、有利には、次のスピーチ・パケットが再生される前に、ジッタ保護をその最大に構築するために挿入される。

10

【0033】

同様に、図4に示す本発明の説明的な実施形態によれば、トーク・スパートは、有利には、ハーフレートが受信される時に終了すると仮定される。(EVRコードブックにおいて、ハーフレート・フレームは、他の状況でも起こる可能性があるが、トーク・スパートは、必ず、こうしたハーフレート・パケットで終了することに留意されたい。) そのため、ハーフレート・パケットが受信されると、全体の仮想ジッタ・バッファ・コンテンツは、パケットの基準再生時間 (たとえば、EVRコードブックについて20ミリ秒) に圧縮される。現在のトーク・スパートの最後の音声パケットが再生される時まで、仮想ジッタ・バッファ・サイズは、有利には、ゼロ・レーテンシに戻り、このことは、エンコーダとデコーダの間の時間ラグが、現在のトーク・スパートが始まった時に時間ラグがそうであったところに戻ることを意味する。これは、有利には、送信器のトーク・スパート長が、受信器のトーク・スパート長に等しいことを確実にする。

20

【0034】

特に、図4のフローチャートは、ボックス401にて、カウンタ nS をゼロに設定し、その後、ボックス402にて、次のパケットを取得し、決定ボックス403にて、パケットがフルレート・パケットであるかどうかを判断するために調べることによって始まる。フルレート・パケットでない場合、決定ボックス404は、パケットが1/8レート・パケットかどうかを判断するために調べる。1/8レート・パケットでない場合、フローは、ボックス402に戻り、次のパケットを取得するが、1/8レート・パケットである場合、ボックス405は、カウンタ nS をインクリメントし、その後、決定ボックス406は、最後のフレームが1/2レート・パケットであったかどうかを判断するために調べる。最後のフレームが1/2レート・パケットであった場合、ボックス407は、トーク・スパートの終了を識別し、いずれの場合も (最後のフレームが、1/2レート・パケットであってもなくても)、フローは、ボックス402に戻って、次のパケットを取得する。

30

【0035】

決定ボックス403が、パケットがフルレート・パケットであると判定する場合、決定ボックス408は、最後のフレームが1/8レート・パケットであったかどうかを判断するために調べる。最後のフレームが1/8レート・パケットでなかった場合、決定ボックス409は、最後のフレームが1/2レート・パケットであったかどうかを判断するために調べ、1/2レート・パケットであった場合、ボックス410は、トーク・スパートが続いていることを識別する。その後、最後のフレームが、1/8レート・パケットであったか否かにかかわらず、フローは、ボックス401に戻って、カウンタ nS をゼロにリセットする。

40

【0036】

一方、決定ボックス408が、最後のフレームが1/8レート・パケットであったと判定する場合、決定ボックス411は、カウンタ nS の値が、閾値 $nThr0$ と $nThr1$ の間にあるかどうかを判断するために調べる。カウンタ nS の値が、これらの閾値の間に

50

ある場合、ボックス412は、新しいトーク・スパートの開始を識別する。一方、カウンタnSの値が、これらの閾値の間でない場合、ボックス413は、新しい会話ターン内の新しいトーク・スパートの開始または語の間の無音を識別する。いずれの場合も(カウンタnSの値がこれらの閾値の間にあってもなくても)、フローは、ブロック401に戻って、カウンタnSをゼロにリセットする。

【0037】

図5は、本発明の説明的な実施形態による、仮想ジッタ・バッファおよび時間スケール修正を使用した、Voice over IP通信ネットワークにおいてネットワーク・ジッタを処理する方法のフロー図を示す。具体的に、図5に示す説明的な方法は、有利には、図2に示す、単一トーク・スパートの積極的なトーク・スパートの管理を実施する、時間スケール修正技法を実施する。

10

【0038】

図に示すように、説明的な方法は、ネットワークからの、対応する受信された音声パケットから音声データのフレームのシーケンスを、ジッタ/バッファに記憶し(ブロック501)、トーク・スパートの最初のいくつかのフレームを含む第1サブシーケンスを識別する(ブロック502)。説明的に、このフレーム数は、図2に示すように、4に等しくてもよい。その後、これらのフレームについての、記憶された音声データの時間スケールは、連続して再生出力される(ブロック503)前に、所定の量(説明的に、図2示すように50%)だけ時間伸張される。

【0039】

トーク・スパートの「中央」部分を含む音声データのフレーム(有利には、大多数のパケットを含む)について、これらのフレームについての、記憶された音声データは、連続して再生出力される(ブロック504)。そして、最後に、トーク・スパートの最後のいくつかのフレーム(説明的に、この数は、図2に示すように3に等しくてもよいが、トーク・スパート終了パケットが検出される時に、まだ再生出力されていなかったパケット数に基づく可能性がある)を含む第2サブシーケンスが識別され(ブロック505)、これらのフレームについての、記憶された音声データの時間スケールは、連続して再生出力される(ブロック506)前に、所定の量(具体的は、図2に示すように、3対1の圧縮率)だけ圧縮される。

20

【0040】

[本発明の第2の説明的な実施形態]

本発明の説明的な実施形態によれば、仮想ジッタ・バッファは、有利には、ネットワーク・ジッタが最も問題となる可能性がある、携帯の「ハンドオフ」が起こる時に、ネットワーク・ジッタ保護を提供する。特に、モバイル局が、サービス基地局を、1つの基地局から別の基地局へ切り換える期間は、通常、サービス・ギャップを作成する。本発明のこの説明的な実施形態によれば、このサービス・ギャップは、ハンドオフを予想しての時間伸張およびハンドオフが完了した後の時間圧縮の有利な使用によって埋められる。

30

【0041】

より具体的には、本発明のこの説明的な実施形態によれば、第1基地局から第2基地局への携帯ハンドオフが起ころうとしていると、判断されると、モバイル局(たとえば、セルフォン)は、有利には、モバイル局は、その再生出力の前に、第1基地局から受信する全ての音声パケット(すなわち、ハンドオフ前に受信されたパケット)の時間伸張を実施し始める。それによって、これが、仮想ジッタ・バッファのレーテンシを増加させ、特に、ハンドオフが起こっている期間(すなわち、第2基地局を通した新しい伝送経路が構成されている期間)中に、ネットワーク・ジッタ保護を提供する。モバイル基地局が、第2基地局から音声パケットを受信すると、モバイル基地局は、有利には、第2基地局から受信した全ての音声パケットの時間圧縮を実施し始め、ジッタ・バッファ・レーテンシが、その元の量(たとえば、ゼロ)に戻るまで、その音声パケットを再生出力する。

40

【0042】

本発明の1つの説明的な実施形態によれば、先に述べた本発明の第1の説明的な実施形

50

態のトーク・スパート管理技法、および、ここで述べる本発明の第2の説明的な実施形態の携帯ハンドオフは、有利には、組み合わせられる。特に、先に述べたトーク・スパート管理技法は、トーク・スパート(上述した)内の種々の音声パケットの場所に従ってジッタ・バッファ・レーテンシを付加し、一方、ここで述べる携帯ハンドオフは、有利には、ハンドオフが起ころうとする時に付加的なジッタ・バッファ・レーテンシを増す。そのため、ハンドオフが完了することによって(すなわち、パケットが新しい経路から到着する時)、トーク・スパート管理技法に従って決定された量に戻るまで、時間圧縮が実施される。

【0043】

たとえば、CDMA 2000 EVDO、Rev. A(当業者に熟知されている上述した無線通信規格)において、仮想ジッタ・バッファ内の音声パケットは、有利には、ハンドオフ信号が、DSC(デジタル選択呼出システム(Digital Selective Calling))を介してモバイル局によって送出される時に、ハンドオフにわたって伸張される(所与の時間スケール伸張比だけ時間伸張される)。(携帯ハンドオフを実施するためのDSCチャンネルの使用は、当業者に熟知されている。)その後、パケットが、新しい経路から到着すると、パケットは、有利には、再生がエンコーダ・クロックと同期するまで、たとえば、ジッタ・バッファ・レーテンシが、ゼロか、または、トーク・スパート管理技法に従って決定されたレーテンシに戻ってしまうまで圧縮される。これは、(付加的な)ジッタ保護量が、DSCLEN(ハンドオフが起こるまでの時間量を指示するDSC長、すなわち、モバイル局が第1基地局からパケットを受信し続けることになる時間、および、モバイル局が、その後、第2基地局からパケットを受信し始めることが予想される時間)に比例し、さらに、採用される時間スケール伸張比に比例することを可能にする。

【0044】

図6は、本発明の説明的な実施形態による、仮想ジッタ・バッファの使用によって、携帯ハンドオフ中に、サービス・ギャップを埋める、時間スケール修正技法の動作を示す。図は、基地局Aから基地局Bへと起こる携帯ハンドオフを示し、上部に、個々の音声パケットの受信を、下部に、これらのパケットの結果として得られる再生出力を示す。図に示すように、DSCLENは、4パケット長(たとえば、80ミリ秒)であると仮定され、さらに、時間スケール伸張率は50%であることが仮定される。こうしたパラメータを用いると、40ミリ秒までのサービス・ギャップが、いずれのパケット喪失も起こすことなく、すなわち、サービスが全く妨げられずに起こることができることが、図を見てわかる。(図が、基地局Bからの最初の3つのパケットのほぼ同時の到着を示すことに留意されたい。通常、サービス・ギャップは、新しいネットワーク通信経路を構成するのにかかる時間によって主に、生じるため、こうした発生は一般的である。この経路が構成されてしまうと、伝送されるパケットのいくつかは、ほぼ同時に到着することになる可能性がある。)

【0045】

図7は、図6に示す本発明の説明的な実施形態による、仮想ジッタ・バッファおよび時間スケール修正を使用した、Voice over IP通信ネットワークにおいてネットワーク・ジッタを処理する方法のフロー図を示す。ハンドオフが始まると判定されると、モバイル局は、DSCチャンネルを介してハンドオフ信号を送出する(ボックス701)。その後、各音声パケットからの音声データが、再生出力される準備ができ、パケットが試験されて(決定ボックス702)、パケットが、基地局A(モバイル局にサービス提供してきた元の基地局)から受信されたか、または、基地局B(モバイル局にサービス提供する時に、基地局Aの代わりにする新しい基地局)から受信されたかが判定される。さらに、音声パケットが全く受信されなかったということがよくあり、その場合、従来のパケット喪失隠蔽技法が、有利には、抜けているパケットを「置き換える」ために採用されてもよい。

【0046】

再生出力される音声パケットからの音声データが、基地局 A から受信されたと、判定される場合（決定ボックス 702 によって）、音声データは、本発明の原理に従って時間伸張され（ボックス 703）、その後、再生出力される。一方、再生出力される音声パケットからの音声データが、基地局 B から受信されたと、判定される場合（決定ボックス 702 によって）、音声データは、本発明の原理に従って時間伸縮され（ボックス 704）、その後、再生出力される。そして、パケットが全く受信されない場合、パケットは、従来のパケット喪失隠蔽技法の使用と置き換えられる（ボックス 705）。最後に、再生出力が、エンコーダと同期した（たとえば、ジッタ・バッファがゼロ・レーテンシを有するか、または、トーク・スパート管理も使用される場合、ジッタ・バッファ・レーテンシが、管理に従って決定された量に戻った）と、決定ボックス 706 が判定する場合、ハンド 10
オフ・プロセスは完了し、本発明の説明的な方法は終了する。その他の場合、フローは、決定ボックス 702 に戻って、再生出力される次のパケットの音声データの発信源を調査する。

【0047】

[詳細な説明に対する補遺]

先行する議論の全ては、本発明の一般的な原理を単に説明するのみであることに留意されるべきである。当業者は、本明細書には明示的には述べられないし、示されないが、本発明の原理を具体化し、かつ、本発明の精神および範囲内に含まれる、種々の他の構成を考案することができることが理解されるであろう。たとえば、本発明の上述した説明的な実施形態の全てにおいて、ネットワーク・ジッタ保護の一定の「ベースライン」量が、常 20
に設けられてもよく、その場合、本発明の原理に従って設けられるネットワーク・ジッタ保護が付加されることができ、すなわち、ジッタ・バッファの再生レーテンシは、最初に、ゼロでないサイズに設定することができ、それによって、一定の継続的なネットワーク・ジッタ保護が提供され（一定量の再生レーテンシも提供される）、その後、本発明の技法を使用して、付加的なジッタ・バッファ保護が、有利には、付加的な再生レーテンシを付加することなしで提供される。

【0048】

さらに、本明細書で引用される全ての例および暫定的な言語は、本発明の原理および当技術分野を推進することに対して本発明者によって与えられる概念を読者が理解するのを助けるための、教育的な目的のためだけにあることが特に意図され、こうした具体的に引 30
用される例および条件に限定されることはないと解釈されるべきである。さらに、本発明の原理、態様、および実施形態、ならびに、本発明の特定の例を引用する本明細書の全ての申し立ては、本発明の、構造的等価物と機能的等価物の両方を包含することが意図される。こうした等価物は、現在知られている等価物ならびに将来開発される等価物、すなわち、構造にかかわらず、同じ機能を実施する、任意の被開発要素の両方を含むことも意図される。

【図面の簡単な説明】

【0049】

【図 1】 2 人のスピーカ間の会話ターン、および、結果として得られるスピーカ間の会話遅延のサンプルタイミング図である。 40

【図 2】 本発明の 1 つの説明的な実施形態による、仮想ジッタ・バッファの使用によって単一トーク・スパートの積極的なトーク・スパート管理を実施する時間スケール修正技法の動作を示す図である。

【図 3】 本発明の 1 つの説明的な実施形態による、仮想ジッタ・バッファの使用によって会話ターンの積極的なトーク・スパートの管理を実施する時間スケール修正技法の動作を示す図である。

【図 4】 本発明の説明的な実施形態による、トーク・スパート検出を実施する方法のフロー図である。

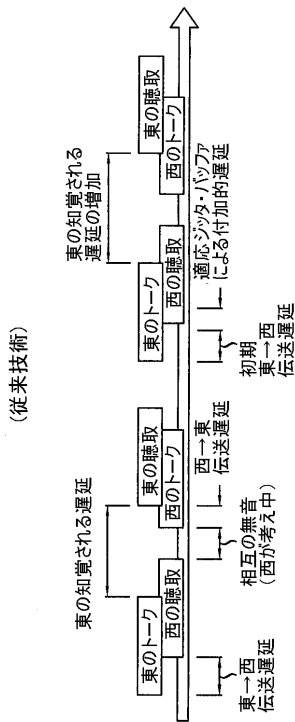
【図 5】 図 2 に示す本発明の説明的な実施形態による、仮想ジッタ・バッファおよび時間スケール修正を使用した、Voice over IP 通信ネットワークにおいてネット 50

ワーク・ジッタを処理する方法のフロー図である。

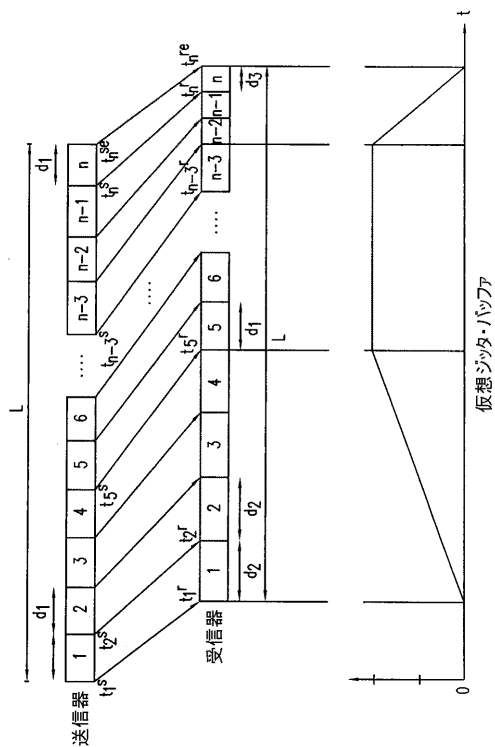
【図6】本発明の説明的な実施形態による、仮想ジッタ・バッファの使用によって、携帯ハンドオフ中に、サービス・ギャップを埋める、時間スケール修正技法の動作を示す図である。

【図7】図6に示す本発明の説明的な実施形態による、仮想ジッタ・バッファおよび時間スケール修正を使用した、Voice over IP通信ネットワークにおいてネットワーク・ジッタを処理する方法のフロー図である。

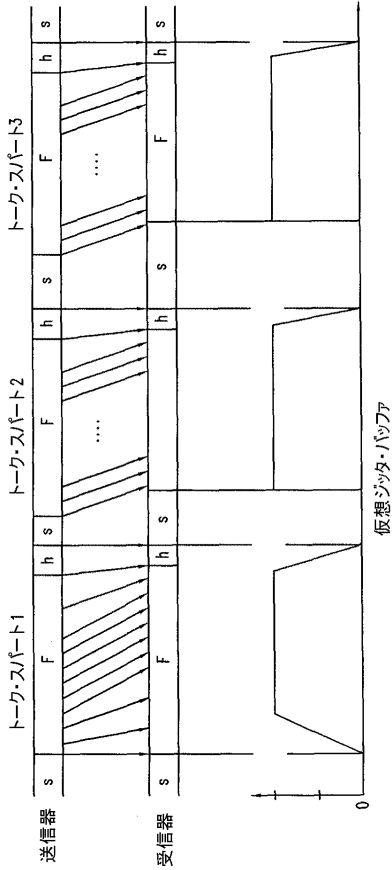
【図1】



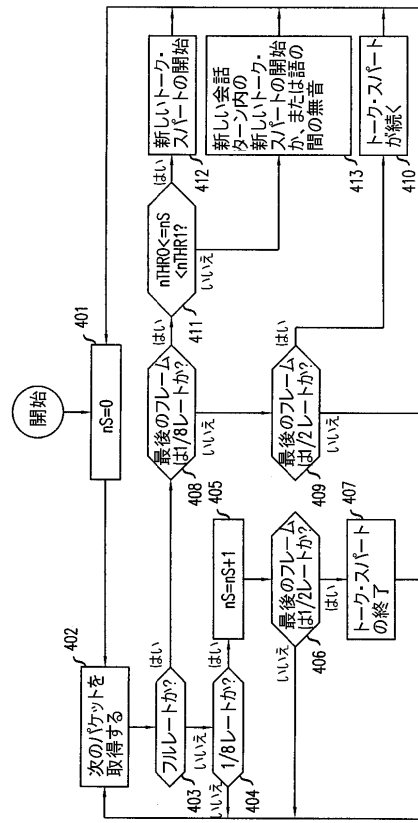
【図2】



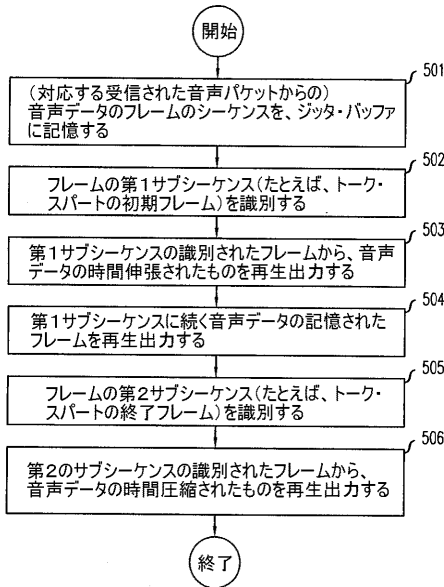
【 図 3 】



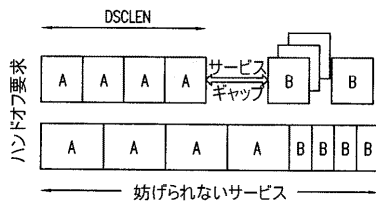
【 図 4 】



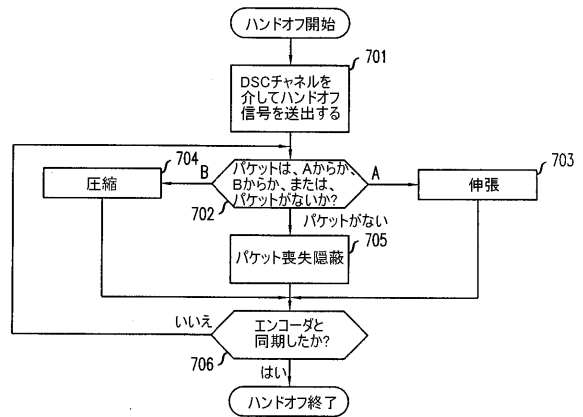
【 図 5 】



【 図 6 】



【 図 7 】



フロントページの続き

- (74)代理人 100101498
弁理士 越智 隆夫
- (74)代理人 100096688
弁理士 本宮 照久
- (74)代理人 100104352
弁理士 朝日 伸光
- (74)代理人 100128657
弁理士 三山 勝巳
- (72)発明者 ミンキュー リー
アメリカ合衆国 0 8 5 5 1 ニュージャーシィ, リンゴーズ, ホワイトハウス ウェイ 3
- (72)発明者 ジェームズ ウィリアム マックゴワン
アメリカ合衆国 0 8 8 8 9 ニュージャーシィ, ホワイトハウス ステーション, ストロベリー
コート 5
- (72)発明者 マイケル チャールズ レッチオーネ
アメリカ合衆国 0 7 1 1 0 ニュージャーシィ, ナトレイ, パサイック アヴェニュー 5 6 5
- Fターム(参考) 5K030 HA08 HB01 JA07 JT01 KA03 MB12 MB13
5K201 AA01 CA02 DB00 EA04 EA07