

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3898660号
(P3898660)

(45) 発行日 平成19年3月28日(2007.3.28)

(24) 登録日 平成19年1月5日(2007.1.5)

(51) Int. Cl.

F I

G 1 0 L 15/00	(2006.01)	G 1 0 L 15/00	2 0 0 A
G 1 0 L 15/14	(2006.01)	G 1 0 L 15/14	2 0 0 Z
G 1 0 L 15/18	(2006.01)	G 1 0 L 15/18	2 0 0 Z
G 1 0 L 13/00	(2006.01)	G 1 0 L 13/00	1 0 0 N
G 1 0 L 13/02	(2006.01)	G 1 0 L 13/02	1 3 0 C

請求項の数 7 (全 11 頁)

(21) 出願番号 特願2003-65832 (P2003-65832)
 (22) 出願日 平成15年3月12日(2003.3.12)
 (65) 公開番号 特開2004-272128 (P2004-272128A)
 (43) 公開日 平成16年9月30日(2004.9.30)
 審査請求日 平成16年7月7日(2004.7.7)

(出願人による申告) 国などの委託研究の成果に係る特許出願(通信・放送機構、平成14年4月1日付け委託研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」、産業再生法第30条の適用を受けるもの)

(73) 特許権者 393031586
 株式会社国際電気通信基礎技術研究所
 京都府相楽郡精華町光台二丁目2番地2
 (74) 代理人 100099933
 弁理士 清水 敏
 (72) 発明者 黒岩 真吾
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内
 (72) 発明者 遠藤 俊樹
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内
 (72) 発明者 中村 哲
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54) 【発明の名称】 音声信号の復元装置及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

入力される音声信号データ列に消失区間があるか否かを判定し、判定結果を示す第1の信号を出力するための判定手段と、

入力される音声信号データ列に対して、音響モデルと言語モデルとを用いて音声認識を行ない、認識結果を出力するための音声認識手段と、

前記音声認識手段の認識結果から音声合成を行なって音声信号を出力するための音声合成手段と、

前記第1の信号にตอบสนองして、前記入力される音声信号データ列に消失区間がないときは前記入力される音声信号データ列の混合比が大となり、消失区間があるときは前記音声合成手段の出力する音声信号の混合比が大となる様に変化する混合比で、前記入力される音声信号データ列と前記音声合成手段の出力とを混合して出力するための混合手段とを含む、音声信号の復元装置。

【請求項2】

前記音声認識手段は、隠れマルコフモデルを音響モデルとして用いるものであり、

前記音声認識手段は、前記第1の信号にตอบสนองし、前記入力される音声信号データ列の消失区間では、前記隠れマルコフモデルにおける出力尤度を全ての状態において等しいものとして音響モデル尤度を計算する事により音声認識を行なう、請求項1に記載の音声信号の復元装置。

【請求項3】

10

20

前記音声認識手段は、前記入力される音声信号データ列の消失区間の音声を、当該消失区間の直前及び直後の音声信号データ列に基づいて、前記音響モデルと前記言語モデルとを用いて音声認識するための手段を含む、請求項 1 に記載の音声信号の復元装置。

【請求項 4】

前記音声認識手段は、前記入力される音声信号データ列の消失区間の音声を、当該消失区間の直前の音声信号データ列に基づいて、前記音響モデルと前記言語モデルとを用いて音声認識するための手段を含む、請求項 1 に記載の音声信号の復元装置。

【請求項 5】

前記言語モデルは、符号帳の符号の出現に関する言語モデルを含む、請求項 1 ~ 請求項 4 のいずれかに記載の音声信号の復元装置。

10

【請求項 6】

前記混合手段は、前記第 1 の信号にตอบสนองして、前記入力される音声信号データ列に消失区間がないときには前記入力される音声信号データ列を、消失区間があるときは前記音声合成手段の出力する音声信号を、それぞれ選択して出力するための選択手段を含む、請求項 1 ~ 請求項 5 のいずれかに記載の音声信号の復元装置。

【請求項 7】

コンピュータにより実行されると、当該コンピュータを請求項 1 ~ 請求項 6 のいずれかに記載の音声信号の復元装置として動作させる、コンピュータプログラム。

【発明の詳細な説明】

【0001】

20

【発明の属する技術分野】

この発明は、パケットを用いる音声通信に関し、特に、パケットロス等により引き起こされる音声の途切れを解消するための技術に関する。

【0002】

【従来の技術】

従来の回線交換方式の電話に代わり、パケット交換方式の電話が普及しつつある。無線を用いた携帯電話でも、かなりの部分がパケット交換方式となっている。また、IP (Internet Protocol) 電話と呼ばれるインターネット上で使用されるパケット交換方式の電話の普及も著しい。通信のための帯域を使う効率を考えると、近い将来、電話の大部分が回線交換方式からパケット交換方式になるものと予想される。

30

【0003】

パケット交換方式を用いた音声通信で問題となるのは、音声の途切れ（プチプチいう状態の事。「音とび」とも呼ばれる。）である。こうした音声の途切れが生じる原因としては、パケットロス又は無線局の切替え（ハンドオーバ）がある。

【0004】

パケットロスは、ネットワーク上のあるノードにトラヒックが集中して輻輳が生じたり、パケット転送の待ち時間に伴うジッタの吸収に失敗したりしてパケットを廃棄する事により生じる。一般に音声パケットはリアルタイム性が要求されるため、伝送プロトコルとしてUDP (User Datagram Protocol) が用いられ、パケットの再送は行なわれない。そのためこの間の音声信号データは失われる。この様に音声信号データが失われる期間の事を本明細書では消失区間と呼ぶ。

40

【0005】

特にアナログ通信と異なり、デジタル通信では音声の途切れの間には全く相手の音声が聞こえなくなるため、こうした音声の途切れを解消する事が望まれる。

【0006】

こうした問題を解決するために知られている従来の 1 手法は、ITU (国際電気通信連合) により ITU G.711 Appendix 1 として勧告されている手法である。その名称は、「G.711 を用いたパケットロス隠蔽のための高品質でかつ複雑さの低いアルゴリズム (High quality low-complexity algorithm for packet loss concealment with

50

G.711)である。

【0007】

この手法では、パケットロスが発生した時に、その前の信号をピッチ同期で繰返す。すなわちこの手法は、パケットロスが発生した区間は、前の音素が続いている事を前提としている。この手法は、現在いくつかのIP電話で実際に使用されている。

【0008】

【特許文献1】

特開2001-100782号公報

【発明が解決しようとする課題】

従来手法では、直前の音素が続いているという前提であるため、復元する消失区間は最大で6フレーム(60ms)である。それ以上の復元は行なわない。また、原理的に2音素以上の欠落には対応できない。しかし、明らかに、より長い消失区間に対しても音声の復元を行なう様にすることが望ましい。また、2音素以上の欠落であっても復元できるようにすることが望まれる。

10

【0009】

それゆえに本発明の目的は、従来より長い消失区間に対しても音声を復元することができる音声信号の復元装置を提供する事である。

【0010】

この発明の他の目的は、従来より長い消失区間に対しても、かつその区間に2音素以上の欠落があっても、音声信号を復元できる音声信号の復元装置を提供する事である。

20

【0011】

【課題を解決するための手段】

本発明の第1の局面に係る音声信号の復元装置は、入力される音声信号データ列に消失区間があるか否かを判定し、判定結果を示す第1の信号を出力するための判定手段と、入力される音声信号データ列に対して、音響モデルと言語モデルとを用いて音声認識を行ない、認識結果を出力するための音声認識手段と、音声認識手段の認識結果から音声合成を行なって音声信号を出力するための音声合成手段と、第1の信号に回答して変化する混合比で、入力される音声信号データ列と音声合成手段の出力とを混合して出力するための混合手段とを含む。

【0012】

好ましくは、音声認識手段は、隠れマルコフモデルを音響モデルとして用いるものであり、音声認識手段は、第1の信号に回答し、入力される音声信号データ列の消失区間では、隠れマルコフモデルにおける出力尤度を全ての状態において等しいものとして音響モデル尤度を計算する事により音声認識を行なう。

30

【0013】

より好ましくは音声認識手段は、入力される音声信号データ列の消失区間の音声を、当該消失区間の直前及び直後の音声信号データ列に基づいて、音響モデルと言語モデルとを用いて音声認識するための手段を含む。

【0014】

音声認識手段は、入力される音声信号データ列の消失区間の音声を、当該消失区間の直前の音声信号データ列に基づいて、音響モデルと言語モデルとを用いて音声認識するための手段を含んでもよい。

40

【0015】

好ましくは、言語モデルは、符号帳の符号の出現に関する言語モデルを含んでもよい。

【0016】

さらに好ましくは、混合手段は、第1の信号に回答して、入力される音声信号データ列に消失区間がないときは入力される音声信号データ列の混合比が大となり、消失区間があるときは音声合成手段の出力する音声信号の混合比が大となる様に、入力される音声信号データ列と音声合成手段の出力とを混合して出力するための手段を含む。

【0017】

50

混合して出力するための手段は、第1の信号に应答して、入力される音声信号データ列に消失区間がないときには入力される音声信号データ列を、消失区間があるときは音声合成手段の出力する音声信号を、それぞれ選択して出力するための選択手段を含んでもよい。

【0018】

本発明の第2の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを上記したいずれかの音声信号の復元装置として動作させるコンピュータプログラムである。

【0019】

【発明の実施の形態】

従来手法は、信号の欠落区間に関しては、信号処理的な手法を用いてボトムアップに信号の復元又は補間を行ってきた。従来技術の項で説明した例がその典型である。しかし、その様な信号処理的な手法を用いた場合には復元できる信号の長さに限界がある事は明らかである。そこで、全く異なる手法により音声信号を復元する事ができれば従来技術の限界を破る事ができる可能性がある。

10

【0020】

そこで、本実施の形態の音声信号の復元装置では、統計的言語情報及び統計的音韻情報を用いて、トップダウンで音声信号を復元する。より具体的には、特許文献1に紹介されている様な、入力音声の一部に欠落がある場合でもその欠落を含んで音声認識を行なう事ができる音声認識手法を用いて受信音声信号の欠落部の音素片を推定し、この音素片情報より音声合成技術を用いて音声波形を合成し欠落部を復元する。

20

【0021】

- 基本概念 -

以下、本実施の形態の基礎となる欠落部の音声信号の復元方法の基本概念について説明する。

【0022】

【数1】

ある音声信号を

$$X = \{x(1), x(2), \dots, x(i), \dots, x(i+n), \dots, x(N)\}$$

とする。この信号のうち、

$$X_m = \{x(i), \dots, x(i+n)\}$$

がパケットロスにより失われたものとする。このとき、

$$P(\{\tilde{x}(i), \dots, \tilde{x}(i+n)\} | \{x(1), \dots, x(i-1)\}, \{x(i+n+1), \dots, x(N)\}) \quad (1)$$

10

を最大とするような

$$\tilde{X}_m = \{\tilde{x}(i), \dots, \tilde{x}(i+n)\}$$

を求めて下記のように音声を補完する。

$$\{x(1), \dots, x(i-1), \tilde{x}(i), \dots, \tilde{x}(i+n), x(i+n+1), \dots, x(N)\}$$

なお、式(1)の代わりに、

$$P(\{\tilde{x}(i), \dots, \tilde{x}(i+n)\} | \{x(1), \dots, x(i-1)\}) \quad (2)$$

20

を用いて左コンテキストのみから信号を予測することもできる。この場合にはリアルタイム処理が可能となる。また、若干の遅延（ $i+M$ フレーム）を許容することにより、次の確率を最大にする方法も考えられる。

$$P(\{\tilde{x}(i), \dots, \tilde{x}(i+n)\} | \{x(1), \dots, x(i-1)\}, \{x(i+n+1), \dots, x(i+n+M)\}) \quad (3)$$

- 実施方法 -

しかし、基本概念として述べた様に式(1)を直接最大化する事は実際には困難である。そこで本実施の形態の装置では、以下の手順で式(1)を最大化する。

30

【0023】

【数2】

(1) 下記の式を最大化する $\{\tilde{s}(1), \dots, \tilde{s}(L)\}$ を求める。

$$P(\{\tilde{s}(1), \dots, \tilde{s}(L)\} | \{x(1), \dots, x(i-1)\}, \{x(i+n+1), \dots, x(N)\})$$

ただし、 $\tilde{s}(i)$ は、音声信号 $\{x(1), \dots, x(i-1)\}, \{x(i+n+1), \dots, x(N)\}$ が与えられたときの、消失区間 ($i \sim i+n$) を含む音声認識結果 (単語系列、音素系列、状態系列、分布系列のいずれか、ただし $L \leq N$) である。

(2) (1) で求めた $\tilde{s}(i)$ を時刻に対応付ける。通常、音声認識の過程で (1) (2) は同時にこなされる。求めた時系列を $\{\hat{s}(1), \dots, \hat{s}(N)\}$ (以下、「状態時系列」と呼ぶ。) とする。

10

(3) 下記の式を最大にする $\{\tilde{x}(i), \dots, \tilde{x}(i+n)\}$ を求める。

$$P(\{\tilde{x}(i), \dots, \tilde{x}(i+n)\} | \{\hat{s}(1), \dots, \hat{s}(N)\})$$

(4) 以上により、下記のように復元・補完された信号を生成することができる。

$$\{x(1), \dots, x(i-1), \tilde{x}(i), \dots, \tilde{x}(i+n), \dots, x(i+n+1), \dots, x(N)\}$$

20

- 構成 -

図 1 に、本実施の形態に係る音声信号の復元装置のブロック図を示す。この復元装置は、たとえば IP 電話の一部として用いられる。図 1 を参照して、この音声信号の復元装置 30 は、入来する音声パケットの各々から音声信号データ部分を抽出して順次出力する入力部 40 と、入力部 40 から時系列的に出力される音声信号データに対して、音声認識を行なって、認識結果として状態時系列を出力するための音声認識部 46 と、音声認識部 46 が音声認識の際に使用する音声認識用音響モデル 42 及び言語モデル 44 とを含む。音声認識用音響モデル 42 は、音響モデルとして、予め学習済みの多数の HMM (隠れマルコフモデル) を含む。言語モデル 44 は、たとえばある言語コーパスから統計的に得られた、想定されているある自然言語における単語のつながり方に関する知識 (確率モデル) を含む。この言語モデル 44 については、予め当該言語のコーパスから作成しておく。

30

【0024】

音声信号の復元装置 30 はさらに、入来する各パケットを監視し、パケットロスが発生しているか否かを示す付随情報を音声認識部 46 に与えるためのモニタ部 48 と、音声認識部 46 から出力された音素系列に基づいて音声合成を行ない、デジタルの音声信号を出力するための音声合成部 52 と、音声合成部 52 が音声合成の際に用いる音声合成用音響モデル 50 と、入力部 40 からの出力に対して音声認識部 46 での音声認識及び音声合成部 52 での音声合成に要する時間に見合う遅延を与えるためのバッファ 54 と、バッファ 54 の出力及び音声合成部 52 の出力を受け、モニタ部 48 に制御されて、パケットロスが発生していない期間ではバッファ 54 の出力の混合比が大となる様に、またパケットロスが発生している期間では音声合成部 52 の出力の混合比が大となる様に、信号を混合して出力するための信号混合部 56 とを含む。

40

【0025】

信号混合部 56 の出力は、たとえば D/A (Digital-to-Analog) コンバータ 58 に与えられ、図示しない増幅器を経てスピーカ 60 に与えられる。

【0026】

図 2 に入力部 40 に入来する音声パケット 80 の構成を概略的に示す。図 2 を参照して、音声パケット 80 は、ヘッダ 82 と、音声信号データ 86 とを含む。ヘッダ 82 はパケット番号フィールド 84 を含む。モニタ部 48 は、このパケット番号フィールド 84 に格納されているパケット番号を見て、パケットロスがあったか否かを判定する。

50

【 0 0 2 7 】

音声認識部 4 6 が行なう音声認識の詳細については、特許文献 1 の記載と同様である。具体的には、音声認識部 4 6 は、次の様にして音声認識を行なう。

【 0 0 2 8 】

音声認識部 4 6 に音声信号として与えられるのは、パケットから抽出されたデジタルの音声信号データ列（数 1 で示す $x(i)$ など）である。音声認識部 4 6 はこの音声信号データ列を、たとえば MFCC（メル・フリーケンシ・ケプストラム係数）分析を行なうことにより音声のフレームごとの特徴ベクトルに変換する。

【 0 0 2 9 】

音声認識部 4 6 は、こうして得られた個々の特徴ベクトルごとに、特徴ベクトルに対する音響モデルのアーキ毎の尤度を計算する。ここでアーキとは、HMM 音響モデルにおいて、ある状態から次の状態への遷移が可能な経路を示す。

10

【 0 0 3 0 】

HMM 音響モデルにおけるアーキ毎の尤度は、ある状態がある特徴ベクトルを出力する出力確率と、その状態から当該アーキを通して状態遷移が生ずるであろう遷移確率との積で表される。通常は、尤度は、計算の都合上、確率値の対数をとって対数尤度とし、積を和に変えて演算する事が多い。

【 0 0 3 1 】

音声認識部 4 6 はさらに、この様にして HMM 音響モデルにより得られた音響尤度と、言語モデル 4 4 が与える言語尤度とを用いて、前述の(1)～(4)の手順に示された原理により音声認識を行ない、音声信号（音素系列）の復元を行なう。この過程で、特徴ベクトルと、認識結果の音素系列（又は単語系列、状態系列、分布系列）との時間的対応付けも同時に定められる。

20

【 0 0 3 2 】

パケットロスの際の状態遷移では、失われた情報を何らかの仮定を用いて推定する事なしに、残された情報のみで音声認識を行なう。これを「Missing Feature Theory」と呼ぶ。その詳細は特許文献 1 に開示されている。そのために、音声認識部 4 6 は、パケットロスの期間のフレームでは、特徴ベクトルの要素がすべて失われたものとして扱う。すなわち、当該フレームにおける音響モデルの出力尤度は、全ての状態で等しいものとし、音響モデルの遷移確率と、言語尤度とを用いて探索を行なう。

30

【 0 0 3 3 】

- 動作 -

図 1 に示す装置は以下の様に動作する。図 3 に示す様に、音声信号データ列 1 0 0（音声信号データ $x_1 \sim x_9$ など）がパケットとして送信されたが、そのうち音声信号データ $x_4 \sim x_6$ に対応する部分 1 0 2 が失われたものとする。

【 0 0 3 4 】

入力部 4 0 は、これら音声信号データ列を音声認識部 4 6、モニタ部 4 8、及びバッファ 5 4 に与える。モニタ部 4 8 は、入力される音声信号データ列中にパケットロスがあるかを判断し、パケットロスがある場合にはその旨を示す付随情報を音声認識部 4 6 に与える。音声認識部 4 6 は、入力される音声信号データ列に対して常に音声認識用音響モデル 4 2 及び言語モデル 4 4 を用いた音声認識を行ない、結果の状態時系列を音声合成部 5 2 に与える。ただし音声認識部 4 6 は、モニタ部 4 8 からの付随情報によりパケットロスがあると判定されている間は、前述した通り尤度最大の条件で状態遷移を行なう。

40

【 0 0 3 5 】

音声合成部 5 2 は、音声認識部 4 6 から出力される音素系列に基づき、音声合成用音響モデル 5 0 を用いて音声合成を行なってその波形を表す音声信号を信号混合部 5 6 に与える。

【 0 0 3 6 】

バッファ 5 4 は、音声認識部 4 6 による音声認識と音声合成部 5 2 による音声合成に見合う時間だけ音声信号を遅延させて信号混合部 5 6 に与える。

50

【 0 0 3 7 】

モニタ部 4 8 は、パケットロスが生じたときには、そのパケットロスに対応する期間では音声合成部 5 2 の出力の混合比が、それ以外の期間ではバッファ 5 4 の出力の混合比が、それぞれ大きくなる様に信号混合部 5 6 による信号の混合を制御する。典型的には、信号混合部 5 6 はそのパケットロスに対応する期間では音声合成部 5 2 の出力のみを、それ以外の期間ではバッファ 5 4 の出力のみを選択して出力する。混合された音声信号は D / A コンバータ 5 8 に与えられ、アナログ信号に変換されてスピーカ 6 0 により音声に変換される。

【 0 0 3 8 】

この結果、図 3 に示す様にパケットロスがあった期間 1 0 2 は、音声合成部 5 2 の出力 $X_4 \sim X_6$ により補完される。この補完は、音響モデルだけでなく、言語モデルという言語に関する知識を用いて行なわれる。たとえば、1 音節分の音声信号がまるまる欠落してしまっても、前後のコンテキストから統計的にその音節を予測する事が可能である。そして、予測された音節を用いて音声合成を行なって音声の欠落部分を補う。そのため、従来の手法と比較してより長い消失時間に対しても対応する事ができる。また、原理的には、複数の音素が欠落した場合でもそれらを復元する事が可能である。

【 0 0 3 9 】

- コンピュータによる実現 -

本発明に係る音声信号の復元装置 3 0 は、コンピュータハードウェア及びその上で実行されるコンピュータプログラムにより実現できる。そのため、携帯電話又は IP 電話など、実質的にコンピュータハードウェア及びコンピュータプログラムにより動作する装置に組み込んで使用する事ができる。図 4 に、その一例として IP 電話機 1 1 0 のブロック図を示す。

【 0 0 4 0 】

図 4 を参照して、この IP 電話機 1 1 0 は、中央演算処理装置 (CPU) 1 2 0 と、CPU 1 2 0 が接続されたバス 1 2 2 と、いずれもこのバス 1 2 2 に接続された RAM (Random Access Memory) 1 2 4、ROM (Read-Only Memory) 1 3 0、及びネットワークに接続され TCP (Transfer Control Protocol) / IP 通信により音声信号データの packets を送受信するための TCP / IP 通信部 1 2 8 と、バス 1 2 2 に接続されたサウンドボード 1 2 6 と、サウンドボード 1 2 6 に接続されたマイク 1 3 2 及びスピーカ 6 0 とを含む。

【 0 0 4 1 】

上記した音声信号の復元装置 3 0 を実現するコンピュータプログラム及びそのためのデータ (音声認識用音響モデル 4 2、言語モデル 4 4 及び音声合成用音響モデル 5 0 など) は、ROM 1 3 0 に格納される。CPU 1 2 0 が、ROM 1 3 0 からこのコンピュータプログラムの各命令を読み出して実行する事により、上記した音声信号の復元装置 3 0 が実現できる。なお、音声認識部 4 6 としてはモニタ部 4 8 からの付加情報に応じて状態遷移の尤度を変化させる事ができる様にすることを条件として、通常の音声認識プログラムを用いる事ができる。音声合成部 5 2 としても通常の音声合成プログラムを利用できる。

【 0 0 4 2 】

モニタ部 4 8 は、本質的には音声信号データの消失区間を検出し、その区間に応じた付随情報及び信号混合部 5 6 の制御信号を出力できるものであればよい。

【 0 0 4 3 】

なお、この実施の形態では、図 3 に示す様に消失区間の直前及び直後の情報を用いて消失区間を補完した。しかし、本発明はその様な実施の形態に限定されるわけではない。特に時間遅れを避け、リアルタイム処理をする事が望まれる場合には、消失区間の直前区間の音声信号データのみを用いて (消失区間の後の区間の情報を用いずに) 消失区間を含んで音声認識を行なう事もできる。この場合を図 5 に例示する。すなわち、消失区間 1 0 2 に先立つ音声信号である音声信号データ列 $x_1 \sim x_3$ を用い、消失区間も含めて音声認識を行なう事で消失区間の音素片を推定する。その結果である $X_4 \sim X_6$ を用いて消失区間 1 0 6

10

20

30

40

50

を補完する。

【 0 0 4 4 】

また、Mフレーム分の遅延を許容し、前述の式(3)により表される確率を最大にする様に消失区間を推定する様に音声認識部46を構成する事も可能である。

【 0 0 4 5 】

上記した実施の形態の音声信号の復元装置30では、言語モデルとして自然言語を想定している。しかし言語モデルはその様な自然言語に関するものに限定されるわけではない。たとえば、コーデック等で使用される符号帳の各符号を一つの音素と見て、それらに対する統計的情報から言語モデル44を作成する事もできる。また、自然言語の言語モデルを用いる場合であっても、いわゆる単語を単位とするのではなく、統計的なまとまりを持った区間を単語として扱って言語モデルを作成する様にしてもよい。

10

【 0 0 4 6 】

ボトムアップクラスタリングなどの技術により音素片を自動的に分類する事で、既存の音素とは異なる擬似音素片を作成して用いてもよい。

【 0 0 4 7 】

また、本実施の形態では、図1に示すように音声認識用音響モデル42と音響合成用音響モデル50とを別に用意した。しかし本発明はそのような実施の形態に限定されるわけではない。たとえば、音声認識部46と音声合成部52とが使用する音響モデルを共通のものとする事もできる。

【 0 0 4 8 】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内でのすべての変更を含む。

20

【 図面の簡単な説明 】

【 図 1 】 本発明の一実施の形態に係る音声信号の復元装置のブロック図である。

【 図 2 】 音声パケットの構成を示す図である。

【 図 3 】 本発明の一実施の形態の装置の動作を説明するための図である。

【 図 4 】 本発明の一実施の形態の装置を組み込んだIP電話機のブロック図である。

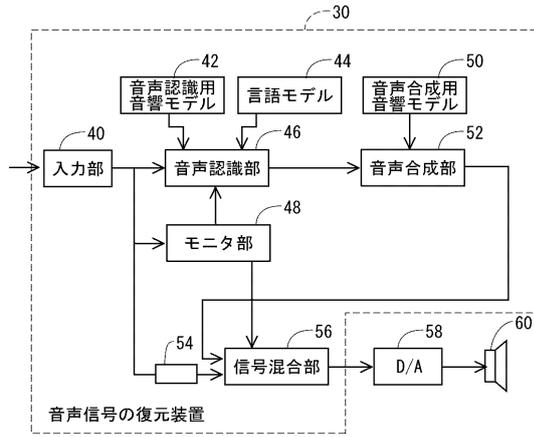
【 図 5 】 本発明の一実施の形態の装置の変形例の動作を説明するための図である。

30

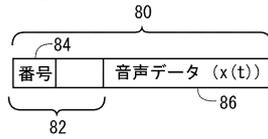
【 符号の説明 】

30 音声信号の復元装置、40 入力部、42 音声認識用音響モデル、44 言語モデル、46 音声認識部、48 モニタ部、50 音声合成用音響モデル、52 音声合成部、54 バッファ、56 信号混合部

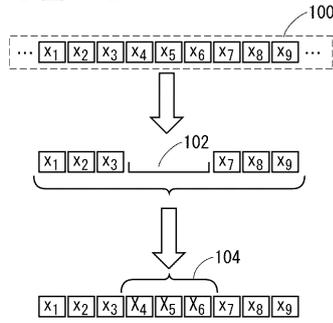
【図1】



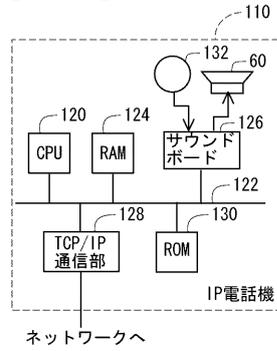
【図2】



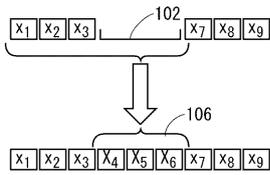
【図3】



【図4】



【図5】



フロントページの続き

審査官 荏原 雄一

(56)参考文献 特開2001-103040(JP,A)
黒岩,他,音声信号の途切れ・オーバーフローへのMissingFeatureTheoryの適用,日本音響学会
講演論文集,1999年,3-Q-16,p.149-150

(58)調査した分野(Int.Cl.,DB名)

G10L 15/00-15/28

G10L 13/00-13/06