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

特許第3552200号

(P3552200)

(45)発行日　平成16年8月11日(2004.8.11)　(24)登録日　平成16年5月14日(2004.5.14)

(51)Int.Cl.[7]		F I		
G10L	19/00	G10L	9/18	A
G10L	13/00	H03M	7/30	B
G10L	13/08	G10L	9/14	S
G10L	15/00	G10L	3/00	H
G10L	15/10	G10L	3/00	R

請求項の数 6　(全 49 頁)　最終頁に続く

(21)出願番号　特願平11-92967
(22)出願日　平成11年3月31日(1999.3.31)
(65)公開番号　特開2000-284799(P2000-284799A)
(43)公開日　平成12年10月13日(2000.10.13)
審査請求日　平成14年2月27日(2002.2.27)

(73)特許権者　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(74)代理人　100058479
弁理士　鈴江　武彦
(74)代理人　100084618
弁理士　村松　貞男
(74)代理人　100068814
弁理士　坪井　淳
(74)代理人　100092196
弁理士　橋本　良郎
(74)代理人　100091351
弁理士　河野　哲
(74)代理人　100088683
弁理士　中村　誠

最終頁に続く

(54)【発明の名称】音声信号伝送装置および音声信号伝送方法

(57)【特許請求の範囲】

【請求項１】

入力音声信号を受けてこれを入力音声情報として出力する音声入力手段と、

該入力音声情報に対して認識処理を施し、その結果を認識結果情報として出力する音声認識手段と、

前記入力音声情報を受けてこれを圧縮処理し、その結果を圧縮結果情報として出力する音声圧縮手段と、

前記認識結果情報および圧縮結果情報を受け、これらのいずれか一方をそれぞれの種別記号と共に選択伝達情報として出力する第１の送信制御手段と、

前記入力音声信号に対する認識処理のスコアが基準以上であり、かつ前記認識結果情報が未学習の音素を含む場合に、該認識結果情報および該圧縮結果情報の組からなる学習対情報を生成し出力する第２の送信制御手段と、

前記選択伝達情報を受け、前記種別記号に基づいて、音声合成手段あるいは音声伸長手段に該選択伝達情報を送る受信制御手段と、

音声合成のための規則あるいは知識である合成規則情報を保持する合成規則記憶手段と、

前記選択伝達情報として受け取った認識結果情報に対し、前記合成規則情報に従って音声合成処理して合成音声出力情報を生成する音声合成手段と、

前記選択伝達情報として受け取った圧縮結果情報に対し、音声信号伸長処理して伸長音声出力情報を出力する音声伸長手段と、

前記合成音声出力情報または前記伸長音声出力情報を受け取り、音声信号として出力する

音声出力手段と、
前記学習対情報を受信したならば、該学習対情報に含まれる前記認識結果情報および前記圧縮結果情報の組に対し音声分析処理、韻律規則学習処理、および音声単位情報学習処理を施すことにより音声合成のための音韻規則および音声単位情報を抽出して新たな合成規則情報を生成し、前記合成規則記憶手段に保持させる学習手段と、
を具備することを特徴とする音声信号伝送装置。
【請求項２】
前記音声認識信号を受け、高精度の音声合成処理を行って高精度合成音声情報を出力する高精度音声合成手段と、
前記音声合成手段により生成された検証用合成音声情報と、前記高精度音声合成手段により生成された高精度合成音声情報とに対し比較検証処理を施し、検証結果を表す検証結果情報を出力する検証手段と、
をさらに具備することを特徴とする請求項１記載の音声信号伝送装置。
【請求項３】
話者毎の特徴を示す話者特徴辞書を有し、入力音声情報に対してこの話者特徴辞書を用いて話者照合処理を行い、話者を特定してこれを話者照合情報として出力する話者照合手段と、
該話者照合情報を受け取り、話者認識結果に応じて、生成する合成音声の種類を変更する合成音声手段と、
を具備することを特徴とする請求項１記載の音声信号伝送装置。
【請求項４】
入力音声信号を受けてこれを入力音声情報として出力する音声入力ステップと、
該入力音声情報に対して認識処理を施し、その結果を認識結果情報として出力する音声認識ステップと、
前記入力音声情報を受けてこれを圧縮処理し、その結果を圧縮結果情報として出力する音声圧縮ステップと、
前記認識結果情報および圧縮結果情報を受け、これらのいずれか一方をそれぞれの種別記号と共に選択伝達情報として出力する第１の送信制御ステップと、
前記入力音声信号に対する認識処理のスコアが基準以上であり、かつ前記認識結果情報が未学習の音素を含む場合に、該認識結果情報および該圧縮結果情報の組からなる学習対情報を生成し出力する第２の送信制御ステップと、
前記選択伝達情報を受け、前記種別記号に基づいて、音声合成手段あるいは音声伸長手段に該選択伝達情報を送る受信制御ステップと、
音声合成のための規則あるいは知識である合成規則情報を保持する合成規則記憶ステップと、
前記選択伝達情報として受け取った認識結果情報に対し、前記合成規則情報に従って音声合成処理して合成音声出力情報を生成する音声合成ステップと、
前記選択伝達情報として受け取った圧縮結果情報に対し、音声信号伸長処理して伸長音声出力情報を出力する音声伸長ステップと、
前記合成音声出力情報または前記伸長音声出力情報を受け取り、音声信号として出力する音声出力ステップと、
前記学習対情報を受信したならば、該学習対情報に含まれる前記認識結果情報および前記圧縮結果情報の組に対し音声分析処理、韻律規則学習処理、および音声単位情報学習処理を施すことにより音声合成のための音韻規則および音声単位情報を抽出して新たな合成規則情報を生成する学習ステップと、
を具備することを特徴とする音声信号伝送方法。
【請求項５】
音声認識信号を受け取り、音声合成ステップに比較してより高精度の音声合成処理を行い、高精度合成音声情報を出力する高精度音声合成ステップと、
該検証用合成音声情報および高精度合成音声情報を受け取り、比較検証処理を施し、検証

結果を表す検証結果情報を出力する検証ステップと、
を具備することを特徴とする請求項４記載の音声信号伝送方法。
【請求項６】
入力音声情報に対して、話者照合処理を行い、話者照合情報として出力する話者照合ステップと、
該話者照合情報を受け取り、出力する送信制御ステップと、
該話者照合情報を受け取り、出力する受信制御制御ステップと、
該話者照合情報受け取り、話者認識結果に応じて、生成する合成音声の種類を変更する合成音声ステップと、
を具備することを特徴とする請求項４記載の音声信号伝送方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、音声信号を通信する音声伝達装置および音声伝達方法に関する。
【０００２】
【従来の技術】
近年、計算機技術および通信技術の進歩により、インターネットに代表されるコンピュータネットワーク網が整備されてきている。また、ＧＵＩ（グラフィカル・ユーザ・インタフェース）技術の進歩などにより、一般ユーザがネットワークに接続されたコンピュータを利用するようになってきている。
【０００３】
このような背景のもと、音声、画像、動画などを含むマルチメディア情報の伝達が活発となってきており、通信の高速化や高能率化、あるいはコスト削減が求められている。
【０００４】
そこで、自動車電話や携帯電話などの移動体通信や企業内通信において、電波の有効利用や通信コストの削減を図る必要があり、これらにおいて、音声信号を低ビットレートで高能率に符号化する技術は重要である。
【０００５】
そして、そのために８［ｋＨｚ］でサンプリングされた電話帯域の音声信号の符号化技術として、これまで様々な方式が提案されている。
【０００６】
例えば、音声波形の振幅値を対数変換して符号化を行うμ‐ｌａｗ　ＰＣＭやＡ‐ｌａｗ　ＰＣＭ、隣接する信号を用いて予測を行い誤差信号を符号化するＡＤＰＣＭ方式などがある。
【０００７】
これらは、音声波形を直接符号化するので波形符号化とも呼ばれ、３２乃至６４［ｋｂｐｓ］でほぼ原音声と差異のない圧縮音声を得ることができる。
【０００８】
これとは別に、２［ｋｂｐｓ］程度の低ビットレートで音声信号を表すことができるパラメトリック符号化がある。
【０００９】
パラメトリック符号化では、音声信号の音韻情報をＬＰＣ係数やＰＡＲＣＯＲ係数などの線形予測係数で構成されるディジタルフィルタで表し、パルス列もしくは白色雑音系列を駆動音源とする。また、波形符号化とパラメトリック符号化のハイブリッド的な方式で、８［ｋｂｐｓ］程度のビットレートで品質の優れた圧縮音声を得ることが可能な分析合成符号化がある。
【００１０】
そして、この分析合成符号化の代表的な手法としては、ＣＥＬＰ（Ｃｏｄｅ　Ｅｘｃｉｔｅｄ　Ｌｉｎｅａｒ　Ｐｒｅｄｉｃｔｉｏｎ）方式が知られている。このＣＥＬＰ方式は、ＡＴ＆Ｔ社ベル研究所のＭ．Ｒ．Ｓｃｈｒｏｅｄｅｒ氏と、Ｂ．Ｓ．Ａｔａｌ氏により

、“Ｃｏｄｅ－Ｅｘｃｉｔｅｄ Ｌｉｎｅａｒ Ｐｒｅｄｉｃｔｉｏｎ（ＣＥＬＰ） Ｈｉｇｈ－Ｑｕａｌｉｔｙ Ｓｐｅｅｃｈ ａｔ Ｖｅｒｙ ｌｏｗ Ｂｉｔ Ｒａｔｅｓ”、Ｐｒｏｃ．ＩＣＡＳＳＰ；１９８５、ｐｐ．９３７－９３９”（文献１）で発表されて以来、高品質な音声が合成できる方式として注目されてきた。

【００１１】

ＣＥＬＰ方式は、短区間音声信号の特徴量をＡｎａｌｙｓｉｓ ｂｙ Ｓｙｎｔｈｅｓｉｓ的手法により、効率よく求めることができる。ここでいう特徴量は、スペクトル包絡を表す短期予測係数、基本周波数を表す長期予測係数、雑音源の種類（Ｃｏｄｅ）およびゲインである。

【００１２】

ＣＥＬＰ方式は、音声波形を直接符号化するのではなく、音声の特徴量を符号化することにより、８［ｋｂｐｓ］以下の非常に小さなビットレートで音声信号を表すことが可能になる。

【００１３】

また、音声認識技術および音声合成技術を手用した知的符号化による情報圧縮および通信方法が提案されている。

【００１４】

＜従来の音声認識＞

ここで従来の音声認識を説明する。現在、音声認識における基本的な方式として、ＨＭＭ（Ｈｉｄｄｅｎ Ｍａｒｋｏｖ Ｍｏｄｅｌ：隠れマルコフモデル）を用いた認識方式がある。

【００１５】

これは、入力音声の各時刻での音響的な特徴を確率事象としてとらえ、各音韻ごとに作成したＨＭＭに基づき、その音響特徴の出現確率を求め、確率値を時間方向に累積してゆき、発声終了時点で最終的に最も高い累積値（＝確率値）を持つカテゴリを認識結果とする方法である。

【００１６】

照合の際には、動的計画法などを用いて効率良く最適な累積経路を求めることが可能である。

【００１７】

＜従来の音声合成＞

一方、従来の音声合成技術について振り返ってみると、当該音声合成技術としては、任意の文章あるいは単語の文字（テキスト）を音声信号に変換する技術の開発が行われてきている。

【００１８】

通常、音声合成システムとしては、言語処理部、音韻処理部、音声波形生成部から構成されている。そして、入力されたテキストは、言語処理部において形態素解析、構文解析、意味解析などを行うことにより、音韻処理で必要な情報、すなわち、語文節の境界、漢字の読み、単語のアクセント、かかり受け、品詞活用形などの言語情報が生成され、次に音韻処理部により、音韻規則に従って、音韻記号列や基本周波数パターンやパワー、継続時間、ポーズの位置などの音韻情報が生成される。

【００１９】

そして、最後に、音声波形生成部において、音韻記号列を合成単位に分解し、合成単位に対応する音源パラメータや伝達特性パラメータなどの音声単位情報を元に、音韻情報に従って制御し、音声信号を生成する。

【００２０】

従来の音声認識技術および音声合成技術としては以上の如きのものがある。

【００２１】

＜知的符号化通信＞

そして、一つの動きとして、以上の如きの音声認識技術および音声合成技術を用いること

によって、送信側で入力された音声信号に対して認識処理を行い、入力された音声信号より通信時のデータ量が小さい、例えば、文字列である認識結果情報や、あるいは音韻列などの中間パラメータ情報を通信路を通して伝送し、受信側で受け取った認識結果情報や中間パラメーダ情報に基づき、合成音声を生成し、出力することにより、従来の符号化による方法より高い圧縮率を実現しようという、知的圧縮通信手法が検討されている。

【００２２】

＜従来方式の問題点＞

ところが、従来の圧縮通信技術には以下に示す問題点がある。

【００２３】

すなわち、従来の圧縮通信技術の問題点は、前述したように従来の圧縮通信技術におけるμ－ｌａｗ　ＰＣＭやＡ－ｌａｗ　ＰＣＭおよびＡＤＰＣＭなどの波形符号化においては、ビットレートが３２～６４［ｋｂｐｓ］のときに非常に高品質な圧縮音声を得ることができるが、ビットレートを低くすると、圧縮音声の品質が急激に劣化するという問題である。

【００２４】

また、パラメトリック符号化では、２［ｋｂｐｓ］程度の低ビットレートでも音声信号を表すことができるが、その圧縮音声の品質は一応の了解性が得られる程度であって品質は十分でなく、パラメトリック符号化のビットレートを高く設定しても、大きな品質の改善は得られないと云う点である。

【００２５】

また、ＣＥＬＰ方式に代表される分析合成符号化では、ビットレートが８［ｋｂｐｓ］程度で十分な品質を得ることができるが、例えば、４［ｋｂｐｓ］程度までビットレートを低下させると、品質劣化が知覚されてしまうという問題である。

【００２６】

つまり、これらのことを考えると、従来の符号化による圧縮通信では、伝達できる音声の品質と圧縮度の間にトレードオフがあるため、可能な圧縮比率に限界があるという大きな問題点があると云うことがわかる。

【００２７】

＜従来の知的符号化技術の問題点＞

一方、知的符号化技術を利用した方法に関しても以下に示す問題点がある。

【００２８】

すなわち、従来の音声認識技術の問題点とは、話者、環境雑音、認識言語、発声様式などの影響により、同じ言葉であっても音声パターンは大きく異なり、認識誤りは避けられないという点である。

【００２９】

＜従来の音声合成技術の問題点＞

また、従来の音声合成技術に関しての問題点としては、各処理部での誤りや、規則の不備、音声単位情報の精度などにより、合成音声が不自然となったり、内容の理解が容易でない、音質が低下するなどの点である。例えば、言語処理部では、形態素解析の誤りによるや読み・アクセント間違い、構文解析、意味解析などの誤りによる、係り受け先の間違いや、品詞文節境界などの間違いなどが考えられる。

【００３０】

これらの間違いにより、音韻処理部などの音韻規則で生成される音韻情報が正しく生成されず、最終的に生成される音声信号での読み上げ内容の相違や不自然さなどへ影響を与える。

【００３１】

また、合成音声では音声単位の接続や継続時間、あるいはピッチの変化などを規則として表現し、入力された文字列の解析結果から音韻規則に基づいて音声を合成するため、アクセント・イントネーションを制御する音韻規則の不備により音質の低下を招く原因となる。

【００３２】
例えば、日本語の平叙文や疑問文などのイントネーションの制御方式として音節ごとの点ピッチ情報を線形に補間して韻律制御する方式“（箱田「文章音声合成におけるピッチパラメータ制御法の検討」日本音響学会音声研究会資料、ＳＰ８８－７　（１９８８））”の検討がなされているが、これは平叙文を考慮した規則リにより生後制御されるため、会話文や案内調などのイントネーションの制御に対しては、対応した規則がないため、不自然となる場合がある。
【００３３】
また、音声単位情報の精度が低いと、生成される合成音声の音質が低下する。
【００３４】
以上のように、従来の知的符号化通信においては、まず音声認識処理に於いては、誤認識の発声を完全に排除することが非常に困難であるため、知的情報圧縮によって通信される情報から、誤りを完全に排除することは非常に困難であり、その情報に基づいて音声合成を行うと、入力された音声信号と相違した誤った音声が出力されるという問題があった。
【００３５】
また、仮に音声認識処理が正しく行われたとしても、上述のように、従来の音声合成技術に関しても、各処理部での誤りや、規則の不備や、あるいは音声単位情報の精度不足などにより、出力される合成音声が、内容の理解が容易でないものとなったり、音質が低下したり、読み上げ内容の相違を発生したり、アクセント・イントネーションの低下などのため不自然となったりするなどのおそれがあるため、知的符号化を用いる通信方法の効果を有用なものとすることが出来なかった。
【００３６】
さらに、音声認識技術と音声合成技術を用いた知的情報圧縮による通信の従来の方式では、単に音声入力の認識結果を元に単に音声合成をおこなうものであるため、音声の持つ声質、あるいはリズムやイントネーションやアクセント、あるいは口調といった音声の持つ個性が伝達の過程で消えてしまうこととなるが、このような個性を反映する情報はできるだけ維持できるようにしたいところである。
【００３７】
【発明が解決しようとする課題】
このように従来は、音声符号化、音声認識、音声合成の点で問題があった。すなわち、音声符号化に関しては音声信号を圧縮処理する際、ビットレートが高い時には非常に高品質な圧縮音声を得ることができるが、ビットレートを低くすると、圧縮音声の品質が急激に劣化するという問題であり、また、パラメトリック符号化では、２［ｋｂｐｓ］程度の低ビットレートでも音声信号を表すことができるが、その圧縮音声の品質は一応の了解が得られる程度でその品質は十分でないという問題と、パラメトリック符号化のビットレートは高く設定しても、大きな品質の改善は得ることができないという問題である。
【００３８】
また、ＣＥＬＰ方式に代表される分析合成符号化では、ビットレートが８［ｋｂｐｓ］程度で十分な品質を得ることができるが、例えば、４［ｋｂｐｓ」程度までビットレートを低下させると品質劣化が知覚されてしまうという問題である。
【００３９】
つまり、従来の符号化による圧縮通信では、伝達できる音声の品質と圧縮度の間にトレードオフがあるため、可能な圧縮比率に限界があるという問題点があることである。
【００４０】
また、音声認識処理に関する問題は、話者、環境雑音、認識語彙、発声様式などの影響により、同じ言葉であっても音声パターンは大きく異なり、認識誤りは避けられないという点である。
【００４１】
また、従来の音声合成技術に関する問題は、各処理部での誤りや、規則の不備、音声単位情報の精度などにより、合成音声が不自然となったり、内容の理解が容易でない、音質が

低下するといった点である。例えば、言語処理部では、形態素解析の誤りによるや読み・アクセント間違い、構文解析、意味解析などの誤りによる、係り受け先の間違いや、品詞文節境界などの間違いなどが考えられ、これらの間違いにより、音韻処理部などの音韻規則で生成される音韻情報が正しく生成されず、最終的に生成される音声信号での読み上げ内容の相違や不自然さなどへ影響を与えるという問題を残していることである。

【００４２】

また、合成音声では音声単位の接続や継続時間、あるいはピッチの変化などを規則として表現し、入力された文字列の解析結果から音韻規則に基づいて音声を合成するため、アクセント・イントネーションを制御する音韻規則の不備により音質の低下を招く原因となるという問題を残している。

【００４３】

また、音声単位情報の精度が低いと、生成される合成音声の音質が低下するという問題を残すことになる。

【００４４】

以上のように、まず音声認識処理に於いては、誤認識の発声を完全に排除することが非常に困難である。そのため、知的情報圧縮によって通信される情報から、誤りを完全に排除することは非常に困難であり、その情報に基づいて音声合成を行うと、入力された音声信号と相違した誤った音声が出力されることになってしまう。

【００４５】

また、仮に音声認識処理が正しく行われたとしても、上述のように、従来の音声合成技術に関しても、各処理部での誤りや、規則の不備や、あるいは音声単位情報の精度不足などにより、出力される合成音声の内容の理解が容易でないものとなったり、音質が低下したり、読み上げ内容の相違を発声したり、アクセント・イントネーションの低下などのため不自然となったりするなどの恐れがあるため、知的符号化を用いる通信方法の効果を有用なものとすることが出来ない。

【００４６】

そこで、この発明の第１の目的とするところは、従来の符号化による圧縮通信では、伝達できる音声の品質と圧縮度の間にトレードオフがあるため、可能な圧縮比率に限界があるという問題点を解決し、より効率の高い音声伝送を実現することにある。

【００４７】

また、この発明の第２の目的とするところは、従来の知的符号化による通信では避けることの出来なかった、音声認識処理での誤認識に起因する伝送内容の誤りを、避けることが出来る音声伝送を実現することにある。

【００４８】

また、この発明の第３の目的とするところは、従来の知的符号化による通信では避けることの出来なかった、音声合成処理での誤生成に起因する伝送内容の誤りを、避けることが出来る音声伝送を実現することにある。

【００４９】

また、以上によって、従来の符号化による圧縮通信より高い効率の伝送を、伝送内容に誤りを起こさずに実現することにある。

【００５０】

【課題を解決するための手段】

上記目的を達成するために、本発明は次のように構成する。

【００５１】

［１］　第一には、入力音声信号を受けてこれを入力音声情報として出力する音声入力手段と、該入力音声情報を受けて認識処理を施し、その結果を認識結果情報として出力する音声認識手段と、該入力音声情報を受けとり、圧縮処理を施し、その結果を圧縮結果情報として出力する音声圧縮手段と、該概認識結果情報および、圧縮結果情報を受けとり、どちらか一方をそれぞれの種別記号と共に、選択伝達情報として出力する送信制御手段と、該選択伝達情報を受けとり、種別記号に基づいて、音声合成手段あるいは音声伸長手段

に、該選択伝達情報を、送る受信制御手段と、該選択伝達情報として受けとった認識結果情報に対して、音声合成処理を行うことによって、合成音声出力情報を生成する音声合成手段と、該選択伝達情報として受けとった圧縮結果情報に対して、音声信号伸長処理を行うことによって、伸長音声出力情報を出力する音声伸長手段と、
該合成音声出力情報および該伸長音声出力情報を受けとり、音声信号として出力する音声出力手段とを具備することを特徴とするものである。

【００５２】

この［１］項に示した構成の本発明にかかる音声伝達装置は、送信部に入力される音声信号に対して、音声認識処理および音声圧縮処理とを施し、認識処理の成否の目標であるスコア情報を参照して、音声認識処理が正しくなされていると考えられる部分については音声認識結果としてのテキストデータを種別記号と共に伝送し、一方、音声認識処理が正しくなされていない可能性があると考えられる部分については、音声圧縮結果を種別記号と共に伝送する。

【００５３】

そして、受信部では、伝送された情報の種別記号に基づいて、音声合成処理あるいは音声伸長処理を実施し、各対応部分の入力された時刻情報にしたがって、音声信号として出力することによって、高い圧縮率で、かつ伝送誤りの無い音声通信を実現するものである。すなわち、音声認識できた語彙については音声に比べてデータ数の遙かに少ないテキストデータ化できるのでテキストデータで送信し、音声認識が旨くいかない部分については音声圧縮したデータを送信することで、トータルとして高い圧縮率で、かつ伝送誤りの無い音声通信を実現する。

【００５４】

［２］　本発明は第二には、音声合成のための規則あるいは知識である合成規則情報を保持する合成規則記憶手段と、該合成規則記憶手段、および音声合成規則を学習するための情報である学習対情報を参照し、合成規則情報を生成し、合成規則記憶手段に保持する学習手段と、選択伝達情報を受けとり、該合成規則記憶手段を参照し、合成音声出力情報を生成する音声合成手段と、認識結果情報、および圧縮結果情報、および学習状況情報記憶手段の少なくとも一つを参照し、認識結果情報と、圧縮結果情報あるいは入力音声信号情報から、学習対情報を生成し出力するか、あるいは選択伝達情報を出力する送信制御手段と、該学習対情報を受けとり、学習状況情報として記録する学習状況記憶手段とを、具備したことを特徴とする。

【００５５】

この［２］項に記載の本発明にかかる音声伝達装置は、上述の［１］項の音声伝送装置の機能に加えてえて以下の機能を持つ。

【００５６】

つまり、送信部に入力される音声信号に対して、認識処理の成否の指標であるスコア情報を参照して、認識処理が正しくなされていると考えられる部分について、学習状況記憶手段を参照し、該認識結果の情報によって受信部において合成音声が正しく生成できるかどうかを調べ、正しく生成できると考えられる場合には、認識結果の情報を伝送し、正しく生成できない可能性があると考えられる場合には、認識結果の情報と圧縮結果の情報の組に基づいて作成した、学習対の情報を識男記号と共に伝送する。

【００５７】

そして、受信部では、学習対の情報が伝達されてきた場合には、該学習対の情報を参照し、学習手段を用いて音声合成のための規則を修正あるいは新規に追加するなどにより学習を行うことによって、入力された音声に対して内容誤りがなくかつ不自然でなく高能率な音声通信や、あるいは声質の一致した高能率な音声通信を実現する。

【００５８】

［３］　本発明は第三には、音声認識信号を受けとり、音声合成手段と同様の音声合成処理を行い、検証用合成音声情報を出力する検証用合成手段と、該検証用合成音声情報および音声入力情報を受けとり、比較検証処理を施し、検証結果を表す検証結果情報を出力

する検証手段と、概検証結果情報を参照して認識結果情報あるいは圧縮結果情報を選択し、選択伝達情報として出力する送信制御手段とを具備したことを特徴とする。

【００５９】

この［３］項に示す構成の本発明にかかる音声伝達装置は、送信部において音声認識結果に基づき、一旦、音声合成処理を行い、入力音声との比較を行い、その結果に応じて、受信部に伝送する信号の種別を決定する。これにより、受信部に於ける合成誤りの発生を未然に防ぎ、より信頼性の高い高能率な音声伝送を実現するものである。

【００６０】

［４］　本発明は第四には、音声認識信号を受けとり、音声合成手段に比較してより高精度の音声合成処理を行い、高精度合成音声情報を出力する高精度音声合成手段と、該検証用合成音声情報および高精度合成音声情報を受けとり、比較検証処理を施し、検証結果を表す検証結果情報を出力する検証手段とを具備したことを特徴とする。

【００６１】

この［４］項記載の本発明にかかる音声伝達装置は、送信部において音声認識結果に基づき、高精度な音声合成処理を行い、また、受信部における音声合成処理と同様の音声合成処理を行い、双方の音声合成処理出力を比較して、その結果に応じて送信部から受信部に伝送する信号の種別を決定する。

【００６２】

これによって、受信部に於ける合成誤りの発生を未然に防ぎ、より信頼性の高い高能率な音声伝送を実現するものである。

【００６３】

すなわち、認識結果を送信した場合に受信部で正しく合成できるかどうかを認識結果の送信前に送信装置側で判定した上で、正しく合成できそうな場合は認識結果を送るようにし、そうでない場合には圧縮結果を送るようにすることにより、受信装置側での合成誤りを抑制する。

【００６４】

これによって、受信部に於ける合成誤りの発生を未然に防ぎ、より信頼性の高い高能率な音声伝送を実現するものである。

【００６５】

［５］　本発明は第五には、入力音声情報に対して、話者照合処理を行い、話者照合情報として出力する話者照合手段と、該話者照合情報を受けとり、出力する送信制御手段と、該話者照合情報を受けとり、出力する受信制御制御手段と、該話者照合情報を受けとり、話者認識結果に応じて、生成する合成音声の種類を変更する合成音声手段を具備したことを特徴とする。

【００６６】

この［５］項の構成による音声伝達装置は、話者照合手段によって、発声者が誰であるかを特定し、話者情報を通信路を介して送し、合成手段において音声に変換する場合に、話者の声に近い音声を出力することを実現する。

【００６７】

［６］　本発明は第六には、入力音声情報に対して音声信号種別判定処理を行い、音声種別情報として出力する音声種別判定手段と、該音声種別情報を参照し、圧縮処理の方法を変更し圧縮結果情報として出力する音声圧縮手段と、選択伝達情報として受けとった該圧縮結果情報の圧縮処理の方法に応じて、音声伸長処理を行う音声伸長手段とを具備することを特徴とする。

【００６８】

この［６］項記載の音声伝達装置は、入力される音声の種別に応じて圧縮処理の種類を変更することにより、高い伝送効率を実現する。

【００６９】

［７］　第七には本発明は、入力音声信号を受けとり入力音声情報として出力する音声入力ステップと、該入力音声情報を受けとり、認識処理を施し、その結果を認識結果情報

として出力する音声認識ステップと、該入力音声情報を受けとり、圧縮処理を施し、その結果を圧縮結果情報として出力する音声圧縮ステップと、該概認識結果情報および、圧縮結果情報を受けとり、どちらか一方をそれぞれの種別記号と共に、選択伝達情報として出力する送信制御ステップと、該選択伝達情報を受けとり、種別記号に基づいて、音声合成ステップあるいは音声伸長ステップに、該選択伝達情報を送る受信制御ステップと、該選択伝達情報として受けとった認識結果情報に対して、音声合成処理を行うことによって、合成音声出力情報を生成する音声合成ステップと、該選択伝達情報として受けとった圧縮結果情報に対して、音声信号伸長処理を行うことによって、伸長音声出力情報を出力する音声伸長ステップと、該合成音声出力情報および該伸長音声出力情報を受けとり、音声信号として出力する音声出力ステップとを具備することを特徴とする。

【００７０】

本発明にかかる音声伝達方法は、装置として実現する上述の第一の構成と、同様の効果を実現する。

【００７１】

［８］　第八には本発明は、音声合成のための規則あるいは知識である合成規則情報を保持する合成規則記憶ステップと、該合成規則記憶手段、および音声合成規則を学習するための情報である学習対情報を参照し、合成規則情報を生成し、合成規則記憶手段に保持する学習ステップと、選択伝達情報を受けとり、該合成規則記憶情報を参照し、合成音声出力情報を生成する音声合成ステップと、認識結果情報、および圧縮結果情報、および学習状況情報記憶情報の少なくとも一つを参照し、認識結果情報と、圧縮結果情報あるいは入力音声信号情報から、学習対情報を生成し出力するか、あるいは選択伝達情報を出力する送信制御ステップと、該学習対情報を受けとり、学習状況情報として記録する学習状況記憶ステップとを具備したことを特徴とする。

【００７２】

本発明にかかる音声伝達方法は、装置として実現する上述の第二の構成と、同様の効果を実現する。

【００７３】

［９］　第九には本発明は、音声認識信号を受けとり、音声合成ステップと同様の音声合成処理を行い、検証用合成音声情報を出力する検証用合成ステップと、該検証用合成音声情報および音声入力情報を受けとり、比較検証処理を施し、検証結果を表す検証結果情報を出力する検証ステップと、概検証結果情報を参照して認識結果情報あるいは圧縮結果情報を選択し、選択伝達情報として出力する送信制御ステップとを具備することを特徴とする。

【００７４】

この［９］項の本発明にかかる音声伝達方法は、装置として実現するようにした上記［３］項に示す第三の構成と、同様の効果を実現する方法である。

【００７５】

［１０］　第十には本発明は、音声認識信号を受けとり、音声合成ステップに比較してより高精度の音声合成処理を行い、高精度合成音声情報を出力する高精度音声合成ステップと、該検証用合成音声情報および高精度合成音声情報を受けとり、比較検証処理を施し、検証結果を表す検証結果情報を出力する検証ステップとを、具備したことを特徴とする。

【００７６】

この［１０］項の本発明にかかる音声伝達方法は、装置として実現するようにした上記［４］項に示す第四の構成と、同様の効果を実現する方法である。

【００７７】

［１１］　第十一には本発明は、入力音声情報に対して、話者照合処理を行い、話者照合情報として出力する話者照合ステップと、該話者照合情報を受けて出力する送信制御ステップと、該話者照合情報を受けとり、出力する受信制御制御ステップと、該話者照合情報受けとり、話者認識結果に応じて、生成する合成音声の種類を変更する合成音声ステッ

プとを具備したことを特徴とする。
【００７８】
この［１１］項の本発明にかかる音声伝達方法は、装置として実現するようにした上記［５］項に示す第五の構成と、同様の効果を実現する方法である。
【００７９】
［１２］　第十二には本発明は、入力音声情報に対して、音声信号種別判定処理を行い、音声種別情報として出力する音声種別判定ステップと、該音声種情報を参照し、圧縮処理の方法を変更し圧縮結果情報として出力する音声圧縮ステップと、選択伝達情報として受けとった該圧縮結果情報の圧縮処理の方法に応じて、音声伸長処理を行う音声伸長ステップとを具備したことを特徴とするものである。
【００８０】
この［１２］項の本発明にかかる音声伝達方法は、装置として実現するようにした上記［６］項に示す第六の構成と、同様の効果を実現する方法である。
【００８１】
【発明の実施の形態】
［第一の実施例］
以下、図面を参照して本発明の第一の実施例に係る音声伝達装置につき説明する。この第一の実施例は基本構成を示しており、上記［１］および［７］に対応する実施例である。この実施例では、音声信号を入力するとこの入力された音声信号を認識処理し、入力音声信号のうちの音声認識処理に成功した語彙の部分については、音声認識した結果としての言葉をテキストデータとして伝達するようにし、これによって高い圧縮率による通信コスト軽減を図るようにし、音声認識処理に失敗した語彙部分や、背景雑音に相当する部分については、音声信号の圧縮処理により通信するようにして、誤りのない通信を実現するものである。
【００８２】
（構成の説明）
図１は、本発明の第一の実施例にかかる音声伝達装置の構成の概要を表しており、図に示すように本装置は、入力部１０１、認識部１０２、圧縮部１０３、送信制御部１０４、受信制御部１０５、合成部１０６、伸長部１０７、および出力部１０８から構成される。
【００８３】
図１における入力部１０１は、マイク、Ａ／Ｄ（アナログ／ディジタル）変換器などによって構成され、利用者から本装置への音声入力信号を受け付け、本装置で処理可能な形式への変換を施し時間情報などと共に、入力音声情報として、認識部１０２および圧縮部１０３などへ出力するように構成している。
【００８４】
図１における認識部１０２は、前記入力部１０１から得られる入力音声情報を受けとり、当該受け取った入力音声情報について認識処理を行い、その認識結果および認識スコア、および本認識部の動作状況などに関する情報のうち少なくとも一つを含む認識結果情報を出力するようにしている。
【００８５】
ここで、音声入力の従来手法に触れておく。
まず、音声入力部１０１において、発声された音声はＡ／Ｄ変換され、ディジタル時系列信号の音声データに変換される。こうして得られた音声データに対して、固定時間（例えば、８ミリ秒：以後この単位をフレームと呼ぶ）ごとに音声がワーを計算し、パワーの時系列を用いて、発声された音声の始終端の時刻を検出する。
【００８６】
この始終端区間内の音声データを分析して認識処理に供する。分析では、例えば高速フーリエ変換などを用いて、周波数分析を行い、時系列信号を周波数がラメータの時系列データに変換する。
【００８７】

例えば、２５６ポイントの高速フーリエ変換を行い、得られた１２８次元のパワースペクトルを、Ｂａｒｋスケールにより、１６次元のバンドパスフィルタ出力に圧縮して、各フレームあたり１６次元の特徴ベクトルからなる音声パターンを生成する。

【００８８】

こうして求められた音声パターンは認識部へ送られ、認識語彙リストに格納されている認識対象語彙について、音声パターンとＨＭＭとのＶｉｔｅｒｂｉ照合により、尤度を計算し、例えば、上位Ｎ位までのスコアを与える語彙と、その語彙を構成する音韻の区間を求める。

【００８９】

＜認識結果の例＞

図４は、認識結果情報の内容の例を表しており、また本図は、認識結果情報が、始端情報Ａ、終端情報Ｂ、認識内容情報Ｃ、およびスコア情報Ｄなどと分類されて記録されるようにしていることも表している。

【００９０】

各認識結果情報において、始端情報Ａおよび終端情報Ｂや、それぞれ対応する入力音声信号区間の開始点および終了点の時刻を記録している。また、認識内容情報Ｃには、対応する入力音声信号の認識処理結果が、例えば、文字列やあるいは音韻列などの形で記録されるようにしている（なお、図４など於いて、記号　“＜”と記号“＞”で囲まれた部分は、その中に記載された文字列に対応する入力音声信号の認識処理結果の情報を表すものとする。）。

【００９１】

また、スコア情報Ｄには、対応する入力音声の認識処理結果の信頼性などを表す尺度である、類似度などのスコア情報が記録されるようにしてある。

【００９２】

図１に示す圧縮部１０３は、前記入力部１０１から得られる入力音声情報を受けとってこれを圧縮処理し、圧縮結果を圧縮結果情報として順次出力するようにしてある。

【００９３】

音声圧縮の従来手法について説明しておく。

圧縮部１０３は、入力部１０１から得られる入力音声情報を受けとり、ＣＥＬＰなどの方法によって、圧縮処理を行い、圧縮結果を圧縮結果情報として出力端子１０２から順次出力する。本実施例では、ＣＥＬＰ符号化を用いて説明するが、本発明はこれに限定されるわけではない。

【００９４】

図２は、ＣＥＬＰ符号化のエンコーダ部の一例としてのブロック構成図を表している。ＣＥＬＰ符号化は、スペクトル包絡符号帳探索部１０３ａ、適応符号帳探索部１０３ｂ、雑音符号帳探索部１０３ｃ、ゲイン符号帳探索部１０３ｄより構成される。各符号帳探索部で探索されたインデックス情報は音声符号化データとして多重化部１０３ｅへ出力される。

【００９５】

次に、音声符号化部１０３の個々の機能について説明する。

【００９６】

スペクトル包絡符号帳探索部１０３ａは成分分離部からの音声信号が主体の成分をフレーム毎に入力し、予め用意しているスペクトル包絡符号帳を探索して、入力された信号のスペクトル包絡をより良く表現することのできる符号帳のインデックスを選択し、このインデックスを多重化部１０３ｅへ出力する。

【００９７】

通常、ＣＥＬＰ方式ではスペクトル包絡を符号化する際に用いるパラメータとしてＬＳＰ（Ｌｉｎｅ　Ｓｐｅｃｔｒｕｍ　Ｐａｉｒ）パラメータを用いるが、これに限られるものではなく、スペクトル包絡を表現できるパラメータであれば他のパラメータも有効である。

【００９８】
適応符号帳探索部１０３ｂは音源の中のピッチ周期で繰り返す成分を表現するために用いる。ＣＥＬＰ方式では、符号化された過去の音源信号を所定の長さだけ適応符号帳として格納し、これを音声符号化部と音声復号化部の両方で持つことにより、指定されたピッチ周期に対応して繰り返す信号を適応符号帳から引き出すことができる構造になっている。
【００９９】
適応符号帳では符号帳からの出力信号とピッチ周期が一対一に対応するため、ピッチ周期を適応符号帳のインデックスに対応させることができる。
【０１００】
このような構造の下、適応符号帳探索部１０３ｂでは、符号帳からの出力信号を合成フィルタで合成したときの合成信号と目標とする音声信号との歪みを聴覚重み付けしたレベルで評価し、その歪みが小さくなるようなピッチ周期を探索する。
【０１０１】
そして、探索されたインデックスを多重化部１０３ｅへ出力する。雑音符号帳探索部１０３ｃは音源の中の雑音的な成分を表現するために用いる。ＣＥＬＰ方式では、音源の雑音成分は雑音符号帳を用いて表される。指定された雑音インデックスに対応して雑音符号帳から様々な雑音信号を引き出すことができる構造になっている。
【０１０２】
このような構造の下、雑音符号帳探索部１０３ｃでは、符号帳からの出力信号を用いて再生される合成音声信号と雑音符号帳探索部５において目標となる音声信号との歪みを聴覚重み付けしたレベルで評価し、その歪みが小さくなるような雑音インデックスを探索する。
【０１０３】
そして、探索された雑音インデックスを多重化部１０３ｅへ出力する。ゲイン符号帳探索部１０３ｄは音源のゲイン成分を表現するために用いる。
【０１０４】
ＣＥＬＰ方式では、ピッチ成分に用いるゲインと雑音成分に用いるゲインの２種類のゲインをゲイン符号帳探索部で符号化する。符号帳探索においては、符号帳から引き出されるゲイン候補を用いて再生される合成音声信号と目標とする音声信号との歪みを聴覚重み付けしたレベルで評価し、その歪みが小さくなるようなゲインインデックスを探索する。
【０１０５】
そして、探索されたゲインインデックスを多重化部１０３ｅへ出力する。多重化部１０３ｅでは、スペクトル包絡のインデックス、適応符号帳のインデックス、雑音符号帳のインデックスおよびゲイン符号帳のインデックスを多重化し、圧縮結果情報として出力端子１０２ｆから順次出力する。
【０１０６】
＜圧縮結果の例＞
図５は、圧縮結果情報の内容の例を表しており、また本図は、圧縮結果情報が、始端情報Ａ、終端情報Ｂ、および圧縮内容清報Ｃなどと分類され記録されるようにしていることも表している。
【０１０７】
各圧縮結果情報において、始端情報Ａおよび終端情報Ｂや、それぞれ対応する入力音声信号区間の開始点および終了点の時刻を記録している。
【０１０８】
また、圧縮内容情報Ｃには、対応する入力音声信号の圧縮処理結果が記録されるようにしている（なお、図５など於いて、記号“（”と記号“）”で囲まれた部分は、その中に記載された文字列に対応する入力音声信号の圧縮処理結果の情報を表すものとする。また、記号“．”は、背景雑音等の無音声区間を表すものとする。）。
【０１０９】
再び図１に戻って構成の説明を続ける。図１における送信制御部１０４は、認識部１０２

から得られる認識結果情報、および圧縮部１０３から得られる圧縮結果情報を受けとり、後述の処理に応じてそのどちらかを選択伝達情報として選択し、通信路を通じて受信制御部１０４へ出力するようにしている。

【０１１０】

なお、本送信制御部１０４は、本発明の効果の実現において重要な役割を担うものであるため、その詳細は後ほど示すこととする。

【０１１１】

＜選択伝達情報の内容の例の説明＞

図６は、送信制御部１０４の出力する選択伝達情報の内容の例を表しており、また、本図は選択伝達情報が種別情報Ａ、および伝達内容情報Ｂなどと分類され、記録されることも表している。各選択伝達情報において、種別情報Ａには、対応する選択伝達情報の種別を表す記号が記録されている。

【０１１２】

ここで、記号「Ｃ」は対応する選択伝達情報が圧縮結果情報であることを表し、また、記号「Ｒ」は対応する選択伝達情報が認識結果情報であることを表している。また、各選択伝達情報において、伝達内容情報Ｂの欄には、種別情報Ａに対応して、認識結果情報あるいは圧縮結果情報が記録されるようにしている。

【０１１３】

再び図１に戻って構成の説明を続ける。図１における受信制御部１０５は、通信路を通じて送信制御部１０４から得られる選択伝達情報を受けとり、種別情報Ａの内容が認識結果情報を表す記号「Ｒ」である場合には、伝達内容情報Ｂの内容を合成部１０６に渡し、種別情報Ａの内容が圧縮結果情報を表す記号「Ｃ」である場合には、伝達内容情報Ｂの内容を、伸長部１０７へ渡す機能を持つ構成としている。

【０１１４】

また、図１における合成部１０６は、受信制御部１０５から認識結果情報を受けとり、以下に示す方法などによって、音声合成処理を行い、合成音声出力情報として出力部１０８へ渡す機能を持つ構成としている。

【０１１５】

＜従来の音声合成処理の概要＞

図７は、合成部１０６の構成を示したものである。合成部１０６は、言語解析部１０６ａ、言語辞書記憶部１０６ｂ、音韻処理部１０６ｃ、音韻規則記憶部１０６ｄ、音声波形生成部１０６ｅ、音声単位情報記憶部１０６ｆで構成される。

【０１１６】

入力されたテキストは、言語解析部１０６ａにより、言語辞書記憶部１０６ｄの情報を参照して形態素解析や構文解析、意味解析を行う事により、音韻処理で必要な言語情報が抽出される。

【０１１７】

例えば、図６の情報の「＜わたしは＞」の部分は、「読み：（わたし）、アクセント：（Ｏ型）、品詞：代名詞／読み：（わ）、アクセント：（Ｏ型）、品詞：（助詞）」などのように解析される。また、始端、終端情報は音韻処理部１０６ｃに送られる。

【０１１８】

次に音韻処理部１０６ｃでは、音韻規則記憶部１０６ｅに格納されている音韻規則に従い、言語情報から、音韻記号列やピッチ周波数パターンやパワー、継続時間、ポーズの位置などの音韻情報を生成する。

【０１１９】

音韻規則は、例えば、読みから対応する音韻記号との対応、読みの並ぴと音韻継続時間との対応、品詞情報、係り受けなどの情報とピッチ周波数パターン・ポーズ位置との対応などを規則として格納している。ここで、音韻の継続時間長を決定する場合、言語処理部１０６ａから送られた始端、終端情報を用いて、時間長を決定する。

【０１２０】

最後に、音声波形生成部１０６ｃは、音韻記号列を合成単位に分解し、音声単位記憶部１０６ｆから、合成単位に対応する音源パラメータや伝達特性パラメータなどの音声単位情報を抽出し、音韻情報にしたがって制御することにより音声信号を生成する。

【０１２１】

＜合成音声出力情報＞

図８は、合成部１０５によって生成され、出力部１０８へ渡される合成音声出力情報の内容の例を示しており、また本図は、合成音声出力情報が、始端情報Ａ、終端情報Ｂ、および合成音声波形情報Ｃなどと分類され記録されることも表している。

【０１２２】

各合成音声出力情報に於いて、始端情報Ａおよび終端情報Ｂは、対応する入力音声区間の始点および終点の時刻を記録しており、また合成音声波形情報Ｃの欄には、対応する合成音声の波形情報が例えばＰＣＭ（Ｐｕｌｓｅ　Ｃｏｄｅ　Ｍｏｄｕｌａｔｉｏｎ）ディジタル波形信号などの形で記録されるようにしている。（なお、図８などに於いて、記号“｛”と記号“｝”で囲まれた部分は、その中に記載された文字列に対応する合成音声の波形情報を表すものとする。

【０１２３】

再び図１に戻って構成の説明を続ける。図１における伸長部１０７は、受信制御部１０５から圧縮結果情報を受けとり、以下の従来方法等による音声伸長処理を行い、音声伸長情報として出力部１０８へ渡す様にしている。

【０１２４】

＜音声伸長処理＞

伸長部１０７は、受信制御部１０５から圧縮結果情報を受けとり、ＣＥＬＰなどの方法によって、音声伸長処理を行い、音声伸長情報として出力部１０８へ渡す。本実施例では、ＣＥＬＰ方式を用いて説明を行うが、本発明はこれに限定されるわけではない。

【０１２５】

図３は、ＣＥＬＰ符号化のデコーダ部のブロック図の例を表している。逆多重化部１０７ｈは符号化部から送られてきた上述した所定の時間単位毎に圧縮結果情報を分離し、スペクトル包絡のインデックス、適応符号帳のインデックス、雑音符号帳のインデックス、ゲインのインデックスを出力する。スペクトル包絡復号部１０７ｄではスペクトル包絡のインデックスと予め用意しているスペクトル包絡符号帳とからスペクトル包絡の情報を再生し、これを合成フィルタ１０７ｆに送る。また、適応音源復号部１０７ａで適応符号帳のインデックスを入力し、これに対応するピッチ周期で繰り返す信号を適応符号帳から引き出し、これを音源再生部１０７ｅに出力する。

【０１２６】

雑音音源復号部１０７ｂは雑音符号帳のインデックスを入力し、これに対応する雑音信号を雑音符号帳から引き出し、これを音源再生部１０７ｅに出力する。

【０１２７】

ゲイン復号部５１０７ｃはゲインのインデックスを入力し、これに対応するピッチ成分に用いるゲインと雑音成分に用いるゲインの２種類のゲインをゲイン符号帳から引き出し、これを音源再生部１０７ｅに出力する。音源再生部１０７ｅは、適応音源復号部１０７ａからのピッチ周期で繰り返す信号（ベクトル）Ｅｐと、雑音音源復号部１０７ｂからの雑音信号（ベクトル）ＥｎとゲインＧｐ，Ｇｎを用いて音源ベクトルＥｘを以下のようにして再生する。

【０１２８】

Ｅｘ＝ＧｐＥｐ＋ＧｎＥｎ

合成フイルタ１０７ｆはスペクトル包絡の情報を用いて音声を合成するための合成フィルタのパラメータを設定し、音源再生部１０７ｅからの音源信号を入力することにより合成音声信号を生成する。

【０１２９】

さらにポストフィルタ１０７ｇで、この合成音声信号に含まれる符号化歪みを整形し、聞

きやすい音となるようにしてから出力部１０８より出力する。
【０１３０】
＜伸長音声出力情報＞
図９は、伸長部１０６によって生成され、出力部１０８へ渡される伸長音声出力情報の内容の例を示しており、かつ、本図９は伸長音声出力情報が始端情報Ａ、終端情報Ｂ、および伸長音声波形情報Ｃなどと分類され記録されることも表している。
【０１３１】
各伸長音声出力情報に於いて、始端情報Ａおよび終端情報Ｂは、対応する入力音声区間の始点および終点の時刻を記録しており、また伸長音声波形情報Ｃの欄には、対応する伸長音声の波形情報が例えば、ＰＣＭ（Ｐｕｌｓｅ　Ｃｏｄｅ　Ｍｏｄｕｌａｔｉｏｎ）デジタル波形信号などの形で記録されるようにしている。（なお、図９などに於いて、記号“｛”と記号“｝”で囲まれた部分は、その中に記載された文字列に対応する伸長音声の波形情報を表すものとする。
【０１３２】
図１における出力部１０８は、例えば、Ｄ／Ａ（デイジタル／アナログ）変換器、あるいはアンプ、あるいはスピーカなどから構成され、合成部１０６から受けとる合成音声情報Ｏｑ、および伸長部１０７から受けとる伸長音声情報Ｏｄを、それぞれに含まれる始端情報Ａの値Ｔａおよび終端情報Ｂの値Ｔｂを参照し、現在時刻Ｔにあらかじめ定めた値Ｈｄを足した時刻（Ｔ＋Ｈｄ）とを比較し、時区間［（Ｔａ＋Ｈｄ），（Ｔｂ＋Ｈｄ）］に、対応するＯｑの合成音声波形情報Ｂの内容Ｗｇあるいは、対応するＯｄの伸長音声波形情報Ｂの内容Ｗｄを、利用者への音声信号として出力する構成としている。
【０１３３】
以上が、本発明の第一の実施例にかかる音声伝達装置を構成する各モジュールの概要であるが、これらの内、入力部１０１、および認識部１０２、および圧縮部１０３、および送信制御部１０４が送信装置を構成し、受信制御部１０５、および合成部１０６、および伸長部１０７、および出力部１０８が受信装置を構成している。そして、両装置を、通信路によって接続することによって、本発明にかかる音声伝達装置が実現される。
【０１３４】
以上が、本発明の第一の実施例にかかる音声伝達装置を構成する各モジュールの概要であるが、これらの内、入力部１０１、および認識部１０２、および圧縮部１０３、および送信制御部１０４が送信装置を構成し、受信制御部１０５、および合成部１０６、および伸長部１０７、および出力部１０８が受信装置を構成し、両装置を、例えば、欧州規格であるＧＳＭや日本規格のＰＤＣ、米国規格の１Ｓ－９５などの無線通信システム、さらにインターネットなどのコンピュータネットワークといった通信路によって接続することにより、本発明にかかる音声伝達装置が実現される。
【０１３５】
つづいて、本発明の効果の実現において主要な役割を担う送信制御部１０４の動作について詳しく説明する。
【０１３６】
なお、ここでは、本装置の本質的な機能および効果と無関係な要因によって、不必要に複雑となることを避けるため、圧縮部１０３おける入力音声情報に対する音声圧縮処理は、認識部１０２における音声認識処理より十分に早く行われ、同じ音声入力に対する、圧縮結果情報は、対応する認識結果情報と、少なくとも同時かあるいは先行して通信制御部１０４に到着するものと仮定する。この仮定とは反対に、認識結果情報が圧縮結果状況より先行して到着する状況や、あるいはその両者の通信制御部１０４への到着順序が変わる状況に対しても同様の効果を実現するよう本装置を構成することも可能であり、上述の仮定によって、一般性を失うことはない。
【０１３７】
図１０は、送信制御部１０４の内部構成の例を示しており、また送信制御部１０４が制御部１０４ａ、および認識情報キュー１０４ａ、および圧縮情報キュー１０４ｂ、および送

信部１０４ｃ、および制御処理部１０４ｄによって構成されていることも示している。
【０１３８】
図１０において１０４ａは、音声認識部１０２から出力される認識情報キュー（Ｆｉｒｓｔ‐ｉｎ‐Ｆｉｒｓｔ‐ｏｕｔ）を表しており、認識結果情報を随時受けとり、図４に例を示した様に各種情報を分類し入力された順序にしたがって随時記録し、制御処理部１０４ｄから入力された順に参照できるようにしている。
【０１３９】
図１０における圧縮情報キュー（Ｆｉｒｓｔ‐ｉｎ‐Ｆｉｒｓｔ‐ｏｕｔ）１０４ｂは、音声圧縮部１０３から出力される圧縮結果情報を随時受けとり、図５に例を示した様に各種情報を分類し、入力される順序にしたがって随時記録し、制御処理部１０４ｄから入力された順に参照できるようにしている。
【０１４０】
また、図１０における送信部１０４ｃは、制御処理部１０４ｄからの制御によって、通信路に選択伝達情報などを出力するようにしている。また、制御処理部１０４ｄは、以下に示す“処理手順Ａ”および“処理手順Ｂ”にしたがった動作をして本装置の構成要素を制御することで、本発明の効果を実現する。
【０１４１】
すなわち、送信部に入力される音声信号に対して、音声認識処理および音声圧縮処理とを施し、認識処理の成否の目標であるスコア情報を参照して、認識処理が正しくなされていると考えられる部分については音声認識結果を種別記号と共に伝送するようにし、一方、認識処理が正しくなされていない可能性があると考えられる部分については、音声圧縮結果を種別記号と共に伝送し、そして、受信部では、伝送された情報の種別記号に基づいて、音声合成処理あるいは音声伸長処理を実施し、各対応部分の入力された時刻情報にしたがって、音声信号として出力することによって、高い圧縮率で、かつ伝送誤りの無い音声通信を実現するものである。すなわち、音声認識できた語彙については音声に比べてデータ数の遙かに少ないテキストデータ化できるのでテキストデータで送信し、音声認識が旨くいかない部分については音声圧縮したデータを送信することで、トータルとして高い圧縮率で、かつ伝送誤りの無い音声通信を実現する。
【０１４２】
なお、図１１は、“処理手順Ａ”を説明するフローチャートである。
【０１４３】
＜“処理手順Ａ”での処理＞
ステップＳ１：　圧縮情報キュー１０４ｂと、認識情報キュー１０４ａと、種別レジスタＶと、内容レジスタｗをクリアし、圧縮結果情報の圧縮情報キュー１０４ｂへの随時記録と、認識結果情報の認識情報キュー１０４ａへの随時記録を開始する。
【０１４４】
ステップＳ２：　認識情報キュー１０４ａが空ならば、ステップＡ２１へ進み、そうでない場合はステップＳ３へ進む。
【０１４５】
ステップＳ３：　認識情報キュー１０４ａの先頭のエントリＲｉを参照し、その始端情報Ａの値Ｔｉｓ、および終端情報Ｂの値Ｔｉｅ、および認識内容情報Ｃの値Ｘｉおよびスコア情報Ｄの値Ｓｉを得る。
【０１４６】
ステップＳ４：　現在の時間Ｔからあらかじめ設定したタイムアウト時間Ｈｔを引いた値（Ｔ－Ｈｔ）と、エントリＲｉの始端の時刻Ｔｉｓを比較し、（Ｔ－Ｈｔ）＜＝Ｔｉｓである場合はステップＡ５へ進み、そうでない場合はステップＳ２０へ進む。
【０１４７】
ステップＳ５：　スコアＳｉとあらかじめ設定した閾値Ｈｓを比較し、ＨＳ＜＝Ｓｉである場合はステップＡ６へ進み、そうでない場合はステップＡ２０へ進む。
【０１４８】

ステップＳ６：　圧縮情報キュー１０４ｂが空である場合はステップＡ２へ進み、そうでない場合はステップＳ７へ進む。

【０１４９】

ステップＳ７：　圧縮情報キュー１０４ｂの先頭のエントリＣｊを参照し、その始点情報Ａの値Ｕｊｓ、および終端情報Ｂの値Ｕｊｅ、および圧縮内容情報Ｙｊを得る。

【０１５０】

ステップＳ８：　エントリＣｊの終端の時刻Ｕｊｅとエントリ Ｒｉの始端の時刻Ｔｉｓを比較し、Ｕｊｅ＜＝Ｔｉｓである場合はステップＳ９へ進み、そうでない場合はステップＳ１１へ進む。

【０１５１】

ステップＳ９：　種別レジスタＶに記号「Ｃ」を記録し、内容レジスタｗにエントリＣｊを複写し、“処理手順Ｂ”に基づく処理を行う。

【０１５２】

ステップＳ１０：　圧縮情報キュー１０４ｂからエントリＣｊを取り除き、ステップＳ６へ進む。

【０１５３】

ステップＳ１１：　エントリＣｊの始端の時刻Ｕｊｓおよび終端の時刻Ｕｊｅと、エントリＲｉの始端の時刻Ｔｊｓを比較し、Ｕｊｓ＜＝Ｔｊｓである場合はステップＳ１２へ進み、そうでない場合はステップＳ１５へ進む。

【０１５４】

ステップＳ１２：　種別レジスタＶに記号「Ｃ」を記録し、内容レジスタｗにエントリＣｊを複写する。

【０１５５】

ステップＳ１３：　内容レジスタｗの圧縮内容情報Ｃの時区間［Ｕｊｓ，Ｔｉｓ］に対応するデータを抽出し、内容レジスタｗの圧縮内容情報Ｃに保存し、内容レジスタｗの終端情報ＢにＴｉｓを記録し、“処理手順Ｂ”に基づく処理を行う。

【０１５６】

ステップＳ１４：　エントリＣｊの圧縮内容情報Ｃの時間Ｔｉｓから終端　（Ｕｊｅ）に対応するデータを抽出し、エントリＣｊの圧縮内容情報Ｃに保存し、エントリＣｊの終端情報ＢにＴｉｓを記録する。

【０１５７】

ステップＳ１５：　種別レジスタＶに記号「Ｒ」を記録し、内容レジスタｗにエントリＲｉを複写し、“処理手順Ｂ”に基づく処理を行う。

【０１５８】

ステップＳ１６：　エントリＲｉの始端の時刻Ｔｉｓ、および終端の時刻Ｔｉｅ、およびエントリＣｊの始端の時刻Ｕｊｓおよび終端の時刻Ｕｊｅを比較し、Ｔｉｓ＜＝Ｕｊｓかつ Ｕｊｅ＜＝Ｔｉｅである場合はステップＳ１７へ進み、そうでない場合はステップＳ１８へ進む。

【０１５９】

ステップＳ１７：　圧縮情報キュー１０４ｂからエントリＣｊを取り除き、ステップＳ１６へ進む。

【０１６０】

ステップＳ１８：　エントリＲｉの始端の時刻Ｔｉｓ、および終端の時刻Ｔｉｅ、およびエントリＣ５の始端の時刻Ｕｊｓおよび終端の時刻Ｕ５ｅを比較し、ＴｉＳ＜＝Ｕｊｓかつ Ｔｉｅ＜＝Ｕｊｅである場合はステップＡ１９へ進み、そうでない場合はステップＳ２０へ進む。

【０１６１】

ステップＳ１９：　エントリＣｊの圧縮内容情報Ｃの時区間［Ｔｉｅ，Ｕｊｅ］分のデータを抽出し、エントリＣｊの圧縮内容情報Ｃに保存し、エントリＣｊの始端情軒ＬにＴｉｅを記録する。

【０１６２】
ステップＳ２０：　　認識情報キュー１０４ａからエントリＲｉを取り除く。
【０１６３】
ステップＳ２１：　　圧縮情報キュー１０４ｂが空ならば、ステップＳ２へ進み、そうでない場合はステップＳ２２へ進む。
【０１６４】
ステップＳ２２：　　現在の時間Ｔから、あらかじめ設定したタイムアウト時間Ｈｔを引いた値（Ｔ－Ｈｔ）と、エントリＣｊの終端の時刻Ｕｊｅを比較し、Ｕｊｅ＜＝（Ｔ－Ｈｔ）である場合はステップＳ２３へ進み、そうでない場合はステップＳ２へ進む。
【０１６５】
ステップＳ２３：　　種別レジスタＶに記号　Ｃ」を記録し、内容レジスタｗに１ントリＣｊを複写し、“処理手順Ｂ”に基づく処理を行う。
【０１６６】
ステップＳ２４：　　圧縮情報キュー１０４ｂからエントリＣｊを取り除き、ステップＳ２２へ進む。
【０１６７】
次に“処理手順Ｂ”を説明する。
【０１６８】
“処理手順Ｂ”
ステップＢ１：　　種別情報として種別レジスタＶの内容を持ち、かつ内容情報として、内容情報レジスタＷの内容を持つ選択伝達情報を、送信部１０４ｃから出力する。
【０１６９】
ステップＢ２：　　種別レジスタＶ、および内容情報レジスタＷをクリアする。
【０１７０】
以上が本装置の構成とその機能である。次に具体的な処理例を説明する。
【０１７１】
＜具体的処理例の説明＞
ここで先ず上述した音声伝達装置の処理について、図を参照して更に詳しく説明する。
【０１７２】
（１）　　まず、“処理手順Ａ”のステップＳ１によって、送信制御部１０４の、圧縮情報キュー１０４ｂと、認識情報キュー１０４ａと、種別レジスタＶと、内容レジスタＷがクリアされ、また圧縮結果情報の圧縮情報キュー１０４ｂへの随時記録と、認識結果情報の認識情報キュー１０４ａへの随時記録が開始される。
【０１７３】
（２）　　ここで、送信側の利用者が「私は佐藤七郎です　（わたしはさとうしちろうです）」と発声し、入力部１０１から入力されたものとする。
【０１７４】
（３）　　この音声入力は入力部１０１によって取り込まれ、音声入力情報とし　て、認識部１０２および圧縮部１０３へと渡される。
【０１７５】
（４）　　この入力音声情報は、認識部１０２で音声認識処理され、図４に示した認識結果情報Ｉ１～Ｉ４に出力され、送信制御部１０４へ随時渡されたものとする。
【０１７６】
（５）　　これら認識結果情報Ｉ１～Ｉ４が、認識情報キュー１０４ａに随時記録される。
【０１７７】
（６）　　また同時に、圧縮部１０３で圧縮処理され、図５に示した圧縮結果情報Ｊ１～Ｊ３が出力され、送信制御部１０４に随時渡されるものとする。
【０１７８】
（７）　　これら圧縮結果情報Ｊ１～Ｊ３が、圧縮結果キュー１０４ｂに随時記録さる。

【０１７９】
ここでは、以上の認識結果情報Ｉ１～Ｉ４および圧縮結果情報Ｊ１～Ｊ３の始端および終端が図１２に示した様な時間的な前後関係を持っているものとし、それぞれの情報はその終端の時刻（Ｔｌｅ，Ｔ２ｅ，Ｔ３ｅ，Ｔ４ｅ，Ｕｌｅ，Ｕ２ｅ，Ｕ３ｅ）に通信制御部１０４に到達するものとする。
【０１８０】
なお、閾値ＨｔおよびＨｄは、認識結果情報１１～１４および圧縮結果情報Ｊ１～Ｊ３に対応する時区間の長さと比較して十分に大きな値が設定されているものとする。また、認識結果情報１１，１２，１４のスコア情報Ｄの値Ｓ１，Ｓ２，Ｓ４は音声認識処理の新表性を判断するための閾値Ｈｓを上回り、認識結果情報１３のスコア情報Ｄの値Ｓ３は閾値ＨＳを下回っていたものとする。
【０１８１】
（ｔ＜Ｔｌｅの場合）
（８）　最初の認識結果情報が到達する時刻Ｔｌｅまでの間は、認識情報キュー１０４ａおよ圧縮情報キュー１０４ｂが双方とも空であるので、“処理手順Ａ”のステップＳ２およびステップＳ２１が構成するループにより、本装置は待機状態となる。
【０１８２】
（ｔ＝Ｔｌｅの場合）
（９）　時刻Ｔｌｅにおいて、認識結果情報Ｉ１が送信制御部１０４へ到達し、認識情報キュー１０４ａに記録される。
【０１８３】
（１０）　ステップＳ２からステップＳ３へ進む。
【０１８４】
（１１）　ステップＳ３で、現在の時刻Ｔから閾値Ｈｔを引いた値と、認識結果Ｉ１の始端の時刻Ｔｌｓが比較されるが、Ｈｔは十分に大きな値であるので、（Ｔ－Ｈｔ）＜＝Ｔｌｓであり、ステップＳ５へ進む。
【０１８５】
（１２）　ステップＳ５で、認識結果情報Ｉ１のスコアＳ１が閾値Ｈｓと比較されるが、ここでは、ＨＳ＜＝Ｓ１であるので、ステップＳ６へ進む。
【０１８６】
（１３）　この時点では、圧縮情報キュー１０４ｂは空であるので、ステップＡ６からステップＳ２へ戻る。
【０１８７】
（Ｔｌｅ＜ｔ＜Ｕｌｅの場合）
以上のステップＡ２～ステップＡ６のループによって、圧縮結果情報Ｊ１が到達するまでの、Ｔｌｅ＜Ｔ＜Ｕ２ｓ（＝Ｕｌｅ）なるＴの間、本装置は待機することになる。
【０１８８】
（ｔ＝Ｕｌｅの場合）
（１４）　時刻Ｕ２ｓ（＝Ｕｌｅ）に於いて、最初の圧縮結果情報Ｊ１が送信制御部１０４にもたらされ、圧縮情報キュー１０４ｂに記録される。
【０１８９】
（１５）　これにより、上記のステップＳ２～Ｓ６のループの中のステップＳ６での分岐条件が成立し、ステップＡ７へ進む。
【０１９０】
（１６）　ステップＳ７において、圧縮結果情報Ｊ１の始端の時刻Ｕ１ｓおよび終端の時刻Ｕｌｅなどを得る。
【０１９１】
（１７）　ステップＳ８において、Ｕｌｅ＜＝Ｔｌｓが成立しないので、ステップＳ１１へ進む。
【０１９２】

（１８）　ステップＳ１１において、Ｕｌｓ＜＝Ｔｌｓが成り立つので、ステップＳ１２へ進む。

【０１９３】

（１９）　ステップＳ１２において、種別レジスタＶに記号「Ｃ」が記録され、内容レジスタｗに１ントリＣｊの内容が複写される。

【０１９４】

（２０）　ステップＳ１３において、内容レジスタｗの圧縮内容〃情報Ｃの時区間［Ｕｌｓ，Ｔｌｓ］に対応するデータが抽出され、内容レジスタｗの圧縮内容情報Ｃに保存され、内容レジスタＷの終端情報ＢにＴｌｓが記録され、“処理手順Ｂ”に基づく処理によって、図６の１Ｄ＝Ｋ１のエントリに示した選択伝達情報が、送信部１０４ｃから送信される。

【０１９５】

以上の処理によって、時区間［Ｕｌｓ，Ｔｌｓ］に対応する音声信号の圧縮結果情報を含む選択伝達情報（図６のＫ１）のエントリが送信されることとなる。

【０１９６】

（２１）　ステップＳ１４において、圧縮結果情報Ｊ１（＝Ｊ１′）の圧縮内容情報Ｃの時区間［Ｔｌｅ，Ｕｌｅ］に対応するデータ「（．．）」が抽出され、圧縮結果情報Ｊ１の圧縮内容情報Ｃに保存され、また終端情報ＢにＴｌｓが記録されることによって、図５の１Ｄ＝Ｊ１の１ントリは、１Ｄ＝σ１′のエントリの如きに書き換えられる。

【０１９７】

（２２）　ステップＳ１５において、種別レジスタＶに記号「Ｒ」が記録され、内容レジスタｗに認識結果情報Ｉ１が記録され、“処理手順Ｂ”に基づく処理によって、送信部から、図６の１Ｄ＝Ｋ２のエントリに示した選択伝達情報が、送信部１０４ｃから送信される。

【０１９８】

以上の処理によって、時区間［Ｔｌｓ，Ｔｌｅ］に対応する音声信号の認識結果情報を含む選択伝達情報（図６のＫ２のエントリ）が送信されることとなる。

【０１９９】

（２３）　ステップＳ１６において、認識結果情報Ｉ１の始端の時刻Ｔｉｓおよび終端の時刻Ｔｌｅ、および圧縮結果情報Ｊ１′の始端の時刻Ｕｌｓおよび終端の時刻Ｕ１ｅが比較され、Ｔｌｓ＜＝Ｔｌｓ、かつ、Ｕｌｅ＜＝Ｔｌｅが成り立たないため、ステップＳ１８へ進む。

【０２００】

（２４）　ステップＳ１８において、認識結果情報Ｉ１の始端の時刻Ｔｌｓ、および終端の時刻Ｔｌｅ、および圧縮結果情報Ｊ１′の始端の時刻Ｕｌｓおよび終端の時刻Ｕ１ｅが比較され、Ｔｌｓ＜＝ＴｌｓかつＴｌｅ＜＝Ｕｌｅが成り立つため、ステップＳ１９へ進む。

【０２０１】

（２５）　ステップＳ１９において、圧縮結果情報Ｊ１′の圧縮内容情報Ｃの時区間［Ｔｌｅ，Ｕｌｅ］分のデータが抽出され、圧縮結果情報Ｊ１′の圧縮内容情報Ｃに保存され、圧縮結果情報Ｊ１′の始端情報ＡにＴｌｅが記録されることによって、図５の１Ｄ＝Ｊ１′のエントリは、１Ｄ＝Ｊ１〃のエントリの様に書き換えられる。

【０２０２】

（２６）　ステップＳ２０において、認識情報キュー１０４ａから、認識結果情報Ｉ１が取り除かれる。

【０２０３】

（２７）　ステップＳ２１において、圧縮情報キュー１０４ｂは空でないので、ステップＳ２２へ進む。

【０２０４】

（２８）　ステップＳ２２において、現在の時間Ｔからあらかじめ設定したタイムアウ

ト時間Ｈｔを引いた値（Ｔ－Ｈｔ）と、圧縮結果情報Ｊｌ″の終端の時刻Ｕｌｅが比較され、その結果、Ｕｌｅ＜＝（Ｔ－Ｈｔ）でないので、ステップＳ２へ進む。

【０２０５】

（２９）　ステップＳ２において、認識情報キュー１０４ａが空であるので、ステップＳ２１へ進む。

【０２０６】

（Ｕ２ｓ＜Ｔ＜Ｔ２ｅの場合）

ここから、上記（２７）～（２８）と同様の処理によって本装置は待機状態となり、次の認識結果情報Ｉ２が到達するまでの、Ｕ２ｓ＜Ｔ＜Ｔ２ｅなるの間、本装置は待機することとなる。

【０２０７】

ｔ＝Ｔ２ｅ

（３０）　時刻Ｔ２ｅにおいて、認識結果情報Ｉ２が送信制御部１０４へ到達し、認識情報キュー１０４ａに記録される。

【０２０８】

（３１）　（１０）～（１２）と同様の処理によって、ステップＳ２から、ステップＳ３、ステップＳ４、およびステップＳ５の処理をへて、ステップＳ６へ進む。

【０２０９】

（３２）　ステップＳ６において、圧縮情報キュー１０４ｂは空でないので、ステップＳ７へ進む。

【０２１０】

（３３）　ステップＳ７において、圧縮情報キュー１０４ｂの先頭のエントリ圧縮結果情報Ｊ１”を参照し、その始点情報Ａの値Ｔｌｅ、および終端情報Ｂの値Ｕｌｅなどを得る。

【０２１１】

（３４）　ステップＳ８において、Ｔｌｅ＜＝Ｔ２ｓであるのでステップＳ９へ進む。

【０２１２】

（３５）　ステップＳ９においては、種別レジスタＶに記号「Ｃ」を記録し、内容レジスタｗに圧縮結果情報Ｊ１″複写し、処理手順Ｂに基づく処理を行うことで、図６のＩＤ＝Ｋ３のエントリに示した選択伝達情報が送信部１０４ｃから送信される。

【０２１３】

以上の処理によって、時区間［Ｔｌｅ，Ｕｌｅ］に対応する音声信号の圧縮結果情報を含む選択伝達情報（図６のＫ３のエントリ）が送信されることとなる。

【０２１４】

（３６）　ステップＳ１０において、圧縮情報キュー１０４ｂから圧縮結果情報Ｊ１”が取り除かれ、ステップＳ６へ進む。

【０２１５】

（３７）　ステップＳ６において、圧縮情報キュー１０４ｂが空である場合なのでステップＳ２へ進む。

【０２１６】

（Ｔ２ｅ＜ｔ＜Ｕ２ｅ）

ここから、（１０）～（１３）と同様の処理によって、ステップＳ２からステップＳ６のループによって、次の圧縮結果情報Ｊ２が到達するまでの、Ｔ２ｅ＜Ｔ＜Ｕ２ｅなるの間、本装置は待機することとなる。

【０２１７】

（ｔ＝Ｕ２ｅ）

（３８）　時刻Ｕ３ｓ（＝Ｕ２ｅ）に於いて、二つめの圧縮結果情報Ｊ２が送信制御部１０４にもたらされ、圧縮情報キュー１０４ｂに記録される。

【０２１８】

（３９）　（１５）～（２０）と同様の処理によって、ステップＳ１３において、内容

レジスタｗの圧縮内容情報Ｃの時区間［Ｕ２ｓ，Ｔ２ｓ］に対応するデータが抽出され、内容レジスタｗの圧縮内容情報Ｃに保存され、内容レジスタｗの終端情報ＢにＴ２ｓが記録され、“処理手順Ｂ”に基づく処理によって、図６の１Ｄ＝Ｋ４のエントリに示した選択伝達情報が、送信部１０４ｃより送信されることとなる。

【０２１９】

以上の処理によって、時区間［Ｕ２ｓ，Ｔ２ｓ］に対応する音声信号の圧縮結果情報を含む選択伝達情報（図６のＫ４のエントリ）が送信されることとなる。

【０２２０】

（４０）　（２１）～（２２）と同様の処理によって、ステップＳ１５において、種別レジスタＶに記号「Ｒ」が記録され、内容レジスタｗに認識結果情報が記録され、“処理手順Ｂ”に基づく処理によって、送信部から、図６の１Ｄ＝Ｋ５のエントリに示した選択伝達情報が、送信部１０４ｃより送信される。

【０２２１】

以上の処理によって、時区間［Ｔ２ｓ，Ｔ２ｅ］に対応する音声信号の認識結果情報を含む選択伝達情報（図６のＫ５のエントリ）が送信されることとなる。

【０２２２】

（４１）　上記（２３）～（２５）と同様の処理によって、ステップＳ１９において、圧縮結果情報Ｊ２′の圧縮内容情報Ｃの時区間［Ｔ２ｅ，Ｕ２ｅ］分のデータが抽出され、圧縮結果情報Ｊ２′の圧縮内容情報Ｃに保存され、圧縮結果情報Ｊ２′の圧縮結果情報Ｊ２′の始端情報ＡにＴ２ｅが記録されることによって、図５の１Ｄ＝Ｊ２′のエントリは、１Ｄ＝Ｊ２″のエントリの如きに書き換えられる。

【０２２３】

（４２）　上記（２６）～（２９）と同様の処理によって、ステップＳ２１へと進む。

【０２２４】

（Ｕ３ｓ＜ｔ＜Ｔ３ｅ）

ここから、上記（２７）～（２８）と同様の処理によって本装置は待機状態となり、次の認識結果情報Ｉ３に到達するまでの、Ｕ３ｓ＜Ｔ＜１３ｅなるＴの間、本装置は待機することとになる。

【０２２５】

ｔ＝Ｔ３ｅ

（４３）　時刻Ｔ３ｅにおいて、認識結果情報Ｉ３が送信制御部１０４へ到達し、認識情報キュー１０４ａに記録される。

【０２２６】

（４４）　上記（１０）～（１１）と同様の処理によって、ステップＳ５へ進む。

【０２２７】

（４５）　ステップＳ５において、認識結果情報Ｉ３のスコアＳ３が閾値Ｈｓと比較されるが、ここでは、Ｈｓ＜＝Ｓ３が成り立たないため、ステップＳ２０へ進む。

【０２２８】

（４６）　ステップＳ２０において、認識情報キュー１０４ａから、認識結果情報Ｉ３が取り除かれる。

【０２２９】

以上の処理によって、信頼性が不十分であると判定された音声認識結果（Ｉ３）は破棄されることとなる。

【０２３０】

（Ｔ３ｅ＜ｔ＜Ｔ４ｅ）

ここから、上記（２７）～（２８）と同様の処理によって本装置は待機状態となり、次の認識結果情報Ｉ４が到達するまでの、Ｔ３ｅ＜Ｔ＜Ｔ４ｅなるＴの間、本装置は待機することとになる。

【０２３１】

ｔ＝Ｔ４ｅ

（４７）　時刻Ｔ４ｅにおいて、認識結果情報Ｉ４が送信制御部１０４へ到達し、認識情報キュー１０４ａに記録される。

【０２３２】

（４８）　上記（３１）～（３５）と同様の処理によって、ステップＳ９において、種別レジスタＶに記号「Ｃ」を記録し、内容レジスタｗに圧縮結果情報Ｊ２″複写し、“処理手順Ｂ”に基づく処理を行うことで、図６のＩＤ＝Ｋ６のエントリに示した選択伝達情報が送信部１０４ｃから送信される。

【０２３３】

以上の処理によって、時区間１Ｔ２ｅ，Ｕ２ｅ］に対応する音声信号の圧縮結果情報を含む選択伝達情報（図６のＫ６のエントリ）が送信されることになる。

【０２３４】

（４９）　上記（３６）～（３７）と同様の処理によってステップＳ２に進む。

【０２３５】

（Ｔ３ｅ＜ｔ＜Ｕ４ｅ）

ここから、上記（１０）～（１３）と同様の処理によってステップＳ２からステップＳ６のループによる次の圧縮結果情報Ｊ３が到着するまでの、Ｔ３ｅ＜ｔ＜Ｕ３ｅなるＴの間、本装置は待機することとなる。

【０２３６】

（ｔ＝Ｕ３ｅ）

（５０）　時刻Ｕ３ｅにおいて、三つ目の圧縮結果情報Ｊ３が送信制御部１０４にもたらされ、圧縮情報キュー１０４ｂに記録される。

【０２３７】

（５１）　上記（１５）～（２０）と同様の処理によって、ステップＳ１３において、内容レジスタｗの圧縮内容惰報Ｃの時区間［Ｕ３ｓ，Ｔ４ｓ］に対応するデータが抽出され、内容レジスタｗの圧縮内容情報Ｃに保存され、内容レジスタｗの終端情報ＢにＴ４ｓが記録され、“処理手順Ｂ”に基づく処理によって、図６のＩＤ＝Ｋ７のエントリに示した選択伝達情報が送信部１０４ｃから送信される。

【０２３８】

以上の処理によって、時区間［Ｕ３ｓ，Ｔ４ｓ］に対応する音声信号の圧縮結果情報を含む選択伝達情報（図６のＫ７のエントリ）が送信されることとなる。（５２）　上記（２１）～（２２）と同様の処理によって、ステップＳ１５において、種別レジスタＶに記号「Ｒ」が記録され、内容レジスタｗに認識結果情報Ｉ４が記録され、“処理手順Ｂ”に基づく処理によって、送信部から図６のＩＤ＝Ｋ８のエントリに示した選択伝達情報が送信部１０４ｃから送信される。

【０２３９】

以上の処理によって、時区間［Ｔ４ｓ，Ｔ４ｅ］に対応する音声信号の認識結果情報を含む選択伝達情報（図６のＫ８のエントリ）が送信されることとなる。

【０２４０】

（５３）　上記（２３）～（２５）と同様の処理によって、ステップＳ１９において、圧縮結果情報Ｊ３′の圧縮内容情報Ｃの時区間［Ｔ４ｅ，Ｕ４ｅ］分のデータが抽出され、圧縮結果情報Ｊ３′の圧縮内容情報Ｃに保存され、圧縮結果情報Ｊ３′の圧縮結果情報Ｊ３′の始端情報ＡにＴ４ｅが記録されることによって、図５のＩＤ＝Ｊ３′のエントリは、ＩＤ＝Ｊ３″のエントリの如きに書き換えられる。

【０２４１】

（５４）　上記（２６）～（２９）と同様の処理によって、ステップＡ２エへと進む。

【０２４２】

（Ｕ３ｅ＜ｔ＜Ｕ３ｅ＋Ｈｔ）

ここから、上記（２７）～（２８）と同様の処理によって本装置は待機状態となり、最後の圧縮結果情報Ｊ３が到達してから、あらかじめ定めたタイムアウト時間Ｈｔが経過するまでの、Ｕ３ｅ＜Ｔ＜Ｕ３ｅ＋ＨｔなるＴの間、ステップＳ２およびステップＳ２１およ

びステップＳ２２が構成するループによって、本装置は待機することとになる。
【０２４３】
（ｔ＝Ｕ３ｅ＋Ｈｔ）
（５５）　時刻Ｔ＝Ｕ３ｅ＋Ｈｔを過ぎた後、ステップＳ２２において、現在の時間Ｔからあらかじめ設定したタイムアウト時間Ｈｔを引いた値（Ｕ３ｅ－Ｈｔ）と、圧縮結果情報Ｊ３″の終端の時刻Ｕ３ｅが比較され、その結果、Ｕ３ｅ＜＝（Ｔ－Ｈｔ）であるので、ステップＳ２３へ進む。
【０２４４】
（５６）　ステップＡ２３において、種別レジスタＶに記号「Ｃ」を記録し、内容レジスタｗに圧縮結果情報Ｊ３″が複写され、“処理手順Ｂ”に基づく処理によって、図６のＩＤ＝Ｋ９のエントリに示した選択伝達惰報が送信部１０４ｃから送信される。
【０２４５】
以上の処理によって時区間［Ｔ４ｅ，Ｕ３ｅ］に対応する音声信号の圧縮結果情報を含む選択伝達情報（図６のＫ７のエントリ）が送信されることとなる。
【０２４６】
（５７）　ステップＳ２４において、圧縮情報キュー１０４ｂから取り除き、ステップＳ２２へ進む。
【０２４７】
ここから本装置は待機状態となる。
【０２４８】
以上に具体例を示した処理によって、本発明の送信制御部１０４から、図６に示した選択伝達情報が受信制御部１０５へ順次伝達される。
【０２４９】
さらに、この選択伝達情報はその種別情報Ａの内容に応じて分配され、音声合成部１０６および音声伸長部１０７で処理されることによって、それぞれから、図８に示した合成音声出力情報および図９に示した伸長音声出力情報が得られ、出力部１０８へもたらされ、これらの情報がそれぞれの時間情報に基づいて、順次出力されることによって、図９のＩＤ＝Ｎ１のエントリ、図８のＩＤ＝Ｍ１のエントリ、図９のＩＤ＝Ｎ２のエントリ、図９のＩＤ＝Ｎ３のエントリ、図８のＩＤ＝Ｍ２のエントリ、図９のＩＤ＝Ｎ４のエントリ、図９のＩＤ＝Ｎ５のエントリ、図８のＩＤ＝Ｍ３のエントリ、および図９のＩＤ＝Ｎ６のエントリの順で出力されることになる結果、
「｛．．｝｛わたしは｝｛．．｝｛．．｝｛さとう｝　｛．．ひち｝｛ろう．．｝｛です｝｛．．｝」
という音声出力がなされる。
【０２５０】
ここでは、本装置に入力された音声信号のうち、音声認識処理に成功した「わたしは」、「さとう」、および「です」の部分については、音声認識結果を伝達することによって、高い圧縮率によって通信コストを軽減することに成功しており、一方、音声認識処理に失敗した「ひちろう」に相当する部分や、背景雑音に相当する部分については、音声圧縮処理による通信が行われることによって、誤りのない通信を実現している。
【０２５１】
すなわち、この実施例では、音声信号を入力するとこの入力された音声信号を認識処理し、入力音声信号のうちの音声認識処理に成功した語彙の部分については、音声認識した結果としての言葉をテキストデータとして伝達するようにし、これによって高い圧縮率による通信コスト軽減を図るようにし、音声認識処理に失敗した語彙部分や、背景雑音に相当する部分については、音声信号の圧縮処理により音声信号の圧縮データを通信するようにして、誤りのない通信を実現するものである。
【０２５２】
従って、このような本システムによれば、従来の符号化による圧縮通信では、伝達できる音声の品質と圧縮度の間にトレードオフがあるため、可能な圧縮比率に限界があったとい

う問題点を解決し、より効率の高い音声伝送が可能となる。また、従来の知的符号化による通信では避けることの出来なかった音声認識処理での誤認識に起因する伝送内容の誤り発生を避けることが出来る音声伝送が可能となる。また、従来の知的符号化による通信では避けることの出来なかった、音声合成処理での誤生成に起因する伝送内容の誤りを、避けることが出来る音声伝送が可能となる。また、以上によって、従来の符号化による圧縮通信より高い効率の伝送で、かつ伝送内容に誤りを起こさない音声伝送が可能となる。

【０２５３】

尚、本発明にかかる音声伝達装置は上述した例に限定されるものではない。例えば、上述の例では、音声入力が利用者からマイクロホンなどによってなされ、また音声出力がスピーカなどによってなされる例が示されているが、入力音声あるいは出力音声として、例えば、磁気記録装置などに記録されている音声データを用いたり、あるいはネットワークを介して得られるデータとしても良い。

【０２５４】

また、あるいは、上述の例では結果認識情報の認識内容情報として、文字列や音韻列などによって、書き下された情報を伝達する例を挙げたが、ここで例えば一意の認識結果ではなく、複数の認識候補をパラメータ情報と共に伝達したり、あるいは認識処理の中間データを伝達するように構成し、音声認識処理における不必要な詳細化を避けて効率化を図るようにすることも可能である。

【０２５５】

またあるいは、上述では、音声認識処理の認識単位が単語である場合の例を示したが、例えば、音韻レベルの認識を行ったり、あるいは文（発話）レベルの認識を行うようにして本発明を構成することも可能である。

【０２５６】

またあるいは、上述の例では、音声信号の開始時点と終了時点の情報を伝送することによって、途切れのない音声出力を実現しているが、さらに入力音声を分析し、基本周波数成分などを分析することによって、入力音声のイントネーション情報などもを送信部から合わせて伝達し、受信部においてこの情報に基づいて合成する合成音声の基本周波数の変化を適宜制御することにより、合成音声波形と、伸長音声波形をイントネーションに関してもスムーズに結合しつつ、元の入力音声の持つイントネーションを保存して伝達できるよう構成することも可能である。

【０２５７】

次に、上記構成のシステムに、学習機能を持たせることによって肉声の持つ個性までも正しく伝達できるようする例を第二の実施例として説明する。

【０２５８】

［第二の実施例］

以下、図面を参照して本発明の第二の実施例に係る音声伝達装置につき説明する。第一の実施例によれば、肉声による発話内容を誤りがなく、かつ、高い圧縮率で伝達できる音声伝達装置が実現されるが、音声信号には、例えば、声質など個性を表す情報が含まれており、これもまた、音声を用いたコミュニケーションで重要な役割を担っている。

【０２５９】

本第二の実施例では、合成音声を生成するための規則などに学習機能を追加することによって、音声の持つ個性までもを正しく伝達でき、かつ高い伝送効率が実現でき、また使用していく過程で学習が進み、次第に伝送効率が向上するといった利点を実現するものである。

【０２６０】

［構成の説明］

図１３は、本発明の第二の実施例にかかる音声伝達装置の構成の概要を表している。図からわかるように、本装置は図１に示した第一の実施例の構成要素に、学習状況記憶部２０９、および学習部２１０および合成規則記憶部２１１を追加した構成としたものである。

【０２６１】

なお、本実施例では、簡単のため、音声認識処理および音声合成処理が音素を最少単位として行われ、かつ、音声認識処理の時区間単位と、音声圧縮処理の時区間単位が同期して行われるものとして説明する。

【０２６２】

図１３において、２０１は入力部であって、第一の実施例での入力部１０１と同様の機能を持つ。また、２０２は認識部であり、第一の実施例での認識部１０２と同様の機能を持ち、入力音声情報に対して音声認識処理を行い、図１４に例を示した如きの認識結果情報を出力する構成としてある。

【０２６３】

すなわち、認識部２０２の出力する認識結果情報は図１４に例示した如く、認識結果情報が、始点情報Ａ、終点情報Ｂ、ラベル情報Ｃ、スコア情報Ｄ、およびＦ０情報Ｅなどと分類されて記録される。

【０２６４】

ここで、始点情報Ａ、終点情報Ｂ、スコア情報Ｄは、図４に示した第一の実施例での認識結果情報の場合と同様である。

【０２６５】

また、ラベル情報Ｃは対応する音声信号の認識結果である音声素片の名前を表しており、またＦ０情報Ｅは、対応する音声信号の基本周波数の代表値などを表しており、これらは、図４に示した第一の実施例での認識内容情報Ｃに相当するものである。（なお、音声素片は、しばしば母音と子音の組合せなどによって表現されるが、ここでは、記号“／”および“／”で囲まれた記述が中に記載された文字に対応する音声素片の識別記号を表現しているものとする。）

図１３に戻って構成を説明する。図１３における構成要素である圧縮部２０３は、第一の実施例での圧縮部１０３と同様の機能を持ち、入力音声情報に対して圧縮処理を行い、図１５に例示した如きの圧縮結果情報を出力する構成としてある。

【０２６６】

圧縮部２０３の出力する図１５に示した圧縮結果情報の例は、圧縮結果情報が、始点情報Ａ、終点情報Ｂ、圧縮内容情報Ｃなどと分類されて記録されることも表している。ここで、始点情報Ａ、終点情報Ｂ、圧縮内容情報Ｃは、図５に示した第一の実施例での圧縮結果情報の場合と同様である。

【０２６７】

図１３に戻って構成を説明する。図１３における構成要素である送信制御部２０４は送信制御を司るものであり、また、受信制御部２０５は受信制御を司るものであって、これら送信制御部２０３の動作、および受信制御部２０４の動作も本実施例の効果の実現にあたって主要な役割を果たしているため、後ほど詳しく説明することとする。

【０２６８】

図１３における合成部２０６は、第一の実施例の合成部１０６と同様の機能を持ち、受信制御部２０４から与えられる選択伝達情報の中の認識内容情報を受け取り、例えば、生成すべき文字列に対応するラベルの系列、あるいは基本周波数の制御信号、あるいは各音素の継続時間の制御信号などを生成し、合成規則記憶部２１１の内容を参照して、ラベル系列の各要素に対して、ラベル名が一致し、かつ基本周波数Ｆ０および継続時間長が最も類似したエントリを集め、例えば、ピッチ同期などの手法を用いて接続したりすることによって音声合成処理を行い、合成音声出力情報を生成し、出力部２０８へ渡す機能を有するものである。

【０２６９】

また、伸長部２０７は、第一の実施例の伸長部１０７と同様の機能を持ち、受信制御部２０４から与えられる選択伝達情報の中の圧縮内容情報を受け取り、音声伸長処理を行い、伸長音声出力情報を生成し、出力部２０８と学習部２１０に渡す機能を有する。

【０２７０】

また、伸長部２０７は、学習部２１０から圧縮された音声素片の情報を受けとり、その伸

長処理結果を学習部２１０に返す機能を有する。また、出力部２０８は、第一の実施例の出力部１０８ど同様の機能を持ち、合成部２０６および、伸長部２０７から、合成音声出力情報および伸長音声出力情報を受け取り、音声出力を行うようにしている。また、学習状況記憶部２０９は、送信部から受信部へ、これまでに送付した学習対情報に関する情報を保持するものであり、ある入力音声の認識結果情報あるいは圧縮結果情報と照合することによって、それらの情報に基づく学習対情報を受信部に送るべきか否かを判断するために利用されるようにしている。

【０２７１】

図１６は、学習状況記憶部２０９の記憶内容の例を表しており、これまでに送付した学習対情報に関する情報が、ラベル情報Ａ、Ｆ０情報Ｂ、継続時間長情報Ｃ、および音声素片情報Ｄなどと分類され記録されるようにしている。

【０２７２】

なお、図１６などで、記号“／”および記号“／”で囲まれた部分は、中に記載された文字に対応する音声の音声素片を表す記号である。

【０２７３】

なお、通常音声素片は、しばしば子音と母音の順序を持った組合せによって表現されることが多いが、ここでは、簡単のため、表層文字で表現することとする。

【０２７４】

学習状況情報記憶部２０９の各エントリにおいて、ラベル情報Ａは、対応する音声素片の識別記号であるラベルが記録される。また、Ｆ０情報Ｂ、および継続時間長情報Ｃの欄には、対応する音声素片の音声合成処理での利用を決定する際の基準となる、基本周波数の代表値、および継続時間長の情報が記録されるようにしている。また、音声素片情報Ｄの欄には、対応する音声素片の波形情報あるいはその圧縮結果などを記録するようにしている。

【０２７５】

図１３における学習部２１０は、受信制御部２０５から受け取る学習対情報を参照して音声合成のための合成規則を生成し、合成規則記憶部２１１に適宜記録する機能を有する。また、合成規則記憶部２１１は、あらかじめ用意した音声合成の為の規則や、学習部２１０から得られる新規あるいは修正された音声合成のための規則を保持する機能を有する。

【０２７６】

図１７は、合成規則記憶部２１１の内容の例を示しているが、図１６に示した、学習状況記憶部２０９の内容の例と同様の情報が記録されることが示されている。

【０２７７】

また、図１８は、合成部２０６および学習部２１０および合成規則記憶部２１１の内部構成の例を表している。ここでは、合成部２０６が、言語処理部２０６ａ、韻律処理部２０６ｂ、および音声波形生成部２０６ｃから構成されることと、合成規則記憶部２１１が、言語辞書記憶部２０６ａ、韻律規則記憶部２１１ｂ、音声単位情報記憶部２１１ｃから構成されることと、学習部２１０が、音声分析部２１０ａ、韻律規則学習部２１０ｂ、および音声単位学習部２０１ｃから構成されることも示されている。

【０２７８】

ここで、言語処理部２０６ａ、韻律処理部２０６ｂ、音声波形生成部２０６ｃ、言語辞書記憶部２１１ａ、韻律規則記憶部２１１ｂ、および音声単位情報記憶部２１１ｃは、それぞれ、図７に示した第一の実施例の言語処理部１０６ａ、韻律処理部１０６ｂ、音声波形生成部１０６ｃ、言語辞書記憶部１０６ｄ、韻律規則記憶部１０６ｅ、および音声単位情報記憶部１０６ｆと同様の働きを持つ。

【０２７９】

学習部２１０では、認識結果と対応する音声信号の伸長された波形信号の組を受け取り、音声分析処理、韻律規則学習処理、および音声単位情報学習処理がなされ、音声合成のための音韻規則および音声単位情報が抽出され、合成規則記憶部２１０に記録するようにしている。

【０２８０】
音声分析部２１０ａは、伸長部２０７から得られる伸長音声出力情報の音声信号を分析して、発声区間中のピッチ周波数パターンや、合成単位の特徴パラメータなどの音声単位情報を抽出する。なお、ここで特徴パラメータとして、ホルマントやケプストラム、ＬＰＣ係数・残差波形対などがある。
【０２８１】
韻律規則学習部２０１ｂは、音声分析部２１０ａで分析した分析情報と認識結果から、韻律規則を新たに学習したり、または既存の韻律規則を修正するようにしている。
【０２８２】
音声単位学習部２０１ｃは、音声分析部２１１ｂの処理で分析して得られた特徴パラメータと合成単位情報記憶部に記憶されている特徴パラメータとを比較して、音声単位情報記憶部２１１ｃの情報を学習、または修正するようにしている。
【０２８３】
＜学習処理＞
ここで、学習部２１０における学習処理の方法の概要について説明する。
従来、音声合成装置において、アクセント・イントネーションなどを制御する音韻規則や、音声波形を生成するための音声単位・情報などは、大量の音声データを学習データとし、学習することにより規則や制御に用いるパラメータ値が適切になるように求めている。
【０２８４】
例として、音韻継続時間長を決定する規則の学習について述べる。
【０２８５】
音韻継続時間長を決定する規則の一つとして、統計的手法を用いて複数の制御要因に対する計算モデルの最適部を数量化１類を用いて行う方式“（酒寄「規則合成のための数量化１類を用いた韻律制御」日本音響学会講演論文集、３－４－１７（１９８６））”がある。
【０２８６】
これは、音韻種類や音韻環境、発話区分長などの制御要因と、計算モデルにおける継続時間長を求める式における制御要素の寄与度を示す係数を、同式により予測された音韻継続時間長と実測された時間長との平均２乗誤差を最小化するように求めるものである。このような手法では、制御要因をもつ音声データが多いほど、適切な係数を得ることができる。
【０２８７】
従って、音声データを多く与えるほど、最適な規則が学習される。また、その他の手法として、回帰木を用いた手法や、出現頻度により最適なものを選択するなどの方法もある。ここでの音韻種類や音韻環境、発話区分長などの制御要因をもつ音声データは、予め用意するか、または、音声認識や音声分析の技術により、所望の制御要因を生成することができる。
【０２８８】
ここでは、規則の学習として音韻継続時間長を決定する規則について述べたが、ピッチパターンや、パワー制御などもの規則も同様な手法で学習することができる。
【０２８９】
次に、音声波形を生成するための特徴パラメータなどの音声単位・情報も、同様に学習データとして与え、学習により最適な音声単位・情報を求めることができる。
【０２９０】
例えば、音声データベースから音声単位を自動的に生成する方法“（中嶋「音韻環境に基づくクラスタリングによる規則合成法」電子情報通信学会論文集Ｄ－ＩＩ、ｖｏｌ、Ｊ７２－Ｄ－ＩＩ、Ｎｏ．８、ｐｐ．１１７４－１１７９（１９８９））”では、スペクトルパラメータの分散に基づいて音韻環境の拘束条件下で音声単位をクラスタリングし、各クラスタのセントロイドを代表の音声単位とするものであり、多くの音声データを与えるほど、最適な音声単位が得られる。

【０２９１】
ここで、音声データベースは、予め用意するか、または、音声認識や音声分析の技術により、音声単位ごとのスペクトルパラメータなどの特徴パラメータを随時生成することができる。
【０２９２】
つづいて、送信制御部２０４の機能について説明する。
【０２９３】
送信制御部２０４は、基本的に第一の実施例の送信制御部１０４と同様の方法によって、同様の機能を実現した上に、以下に示す“処理手順Ｃ”に従った制御によって合成規則の学習のためのデータである学習対情報を送信する機能を実現するものである。
【０２９４】
なお、本送信制御部２０４および受信制御部２０４などでは、音声認識処理の行われれない無音区間については、上述の第一の実施例と同様の処理がなされるものとし、音声認識が行われた部分についてのみ説明を行うこととする。
【０２９５】
図１９は、学習対情報の例を示しており、学習対情報がラベル情報Ａ、Ｆ０情報Ｂ、継続時間情報Ｃ、および音声素片波形情報Ｄなど分類され、記録されていることも示している。なお、これらのは、ラベル情報Ａ、Ｆ０情報Ｂ、継続時間情報Ｃは、図１７に例を示した合成規則記憶部２１１のエントリと同様の意味を持つものとする。また、音声素片波形情報Ｄは、合成規則記憶部２１１のエントリの音声素片情報Ｄの内容に対応する音声素片の波形情報が記録されるようにしている。
【０２９６】
“処理手順Ｃ”の内容を説明する。
【０２９７】
“処理手順Ｃ”
ステップＣ１：　認識部２０２から認識結果情報Ｒｉを受けとる。
【０２９８】
ステップＣ２：　圧縮部２０３から、認識結果Ｒｉと時間的に対応する圧縮結果情報Ｃｊを得る。
【０２９９】
ステップＣ３：　　認識結果情報Ｒｉの始点情報Ａの値Ｔｉｓ′および終点情報Ｂの値Ｔｉｅ′から、式Ｌｉ＝Ｔｉｅ′－Ｔｉｓ′によって、継続時間Ｄｉを得る。
【０３００】
ステップＣ４：　　認識結果情報Ｒｉのスコア情報の値Ｓｉと、予め定めた閾値Ｈｓ′とを比較し、Ｓｉ＞＝Ｈｓ′である場合はステップＣ５へ進み、そうでない場合はステップＣ８へ進む。
【０３０１】
ステップＣ５：　　学習状況記録部２０９を参照し、ラベル情報Ａの値Ｌｘが、認識結果情報Ｒｉのラベル情報Ｃの値Ｌｉと一致し、かつ、Ｆ０情報Ｂの値Ｆｘ２：認識結果情報ＲｉのＦＯ情報Ｅの値Ｆｉの差があらかじめ定めた閾値Ｈｆ以下であり、かつ、継続時間情報Ｃの値と継続時間Ｄｉとの差があらかじめ定めた閾値Ｈｄ以下であり、かつ、その差が最も小さい学習状況記録部２０９のエントリＸを探索し、そのようなエントリＸが存在する場合は、ステップＣ６へ進みそうでない場合はステップＣ７へ進む。
【０３０２】
ステップＣ６：　　ステップＣ５の処理の結果、エントリＸが得られたということは、対応する音声信号を合成部２０６において、合成出力するための合成規則が、合成規則記憶部２１１に既に登録済みであるということを意味するので、認識結果情報を伝達することとし、第一の実施例と同様の方法によって、認識結果Ｒｉに対応する選択伝達情報を送信し、ステップＣ８へ進む。
【０３０３】

ステップＣ７：　一方、ステップＣ５での処理の結果、エントリＸが得られなかったということは、対応する音声信号を合成部２０６において、合成出力するための合成規則が合成規則記憶部２１１に登録されていないことを意味するので、“処理手順Ｄ”によって、対応する音声信号を合成に用いる合成規則を学習部２１０において学習するための情報である学習対情報を送信し、ステップＣ１へ戻る。

【０３０４】

ステップＣ８：　選択伝達情報として、圧縮結果情報を伝達することとし、第一の実施例と同様の方法によって、圧縮結果情報Ｃｊに対応する選択伝達情報を送信し、ステップＣ１へ戻る。

【０３０５】

以上が“処理手順Ｃ”での処理である。

【０３０６】

次に学習対情報を生成する“処理手順Ｄ”の処理内容を説明する。

【０３０７】

“処理手順Ｄ”

ステップＤ１：　認識結果情報Ｒｉのラベル情報Ｃの値Ｌｉ、およびＦ０情報Ｅの値Ｆｉと、継続時間Ｄｉ、および圧縮結果情報Ｃｊの音声素片情報Ｄの値Ｗｊを、それぞれ、ラベル情報Ａ、および、Ｆ０情報Ｂ、および継続時間情報Ｃ、および音声素片情報Ｄとして持つ学習対情報Ｐｋを新たに生成する。

【０３０８】

ステップＤ２：　学習対情報Ｐｋを送信する。

以上が送信制御部２０４の構成と機能である。

続いて、受信制御部２０５の機能について説明する。

【０３０９】

＜受信制御部２０５の機能＞

受信制御部２０５は、第一の実施例の受信制御部１０５と同様の方法によって同様の機能を実現した上で、以下に示す“処理手順Ｅ”に従った制御により、合成規則の学習のためのデータである学習対情報を受信する機能を実現するものである。

【０３１０】

“処理手順Ｅ”を説明する。

【０３１１】

“処理手順Ｅ”ステップＥ１：通信路を通じて送信制御部２０４から選択伝達情報を受けとった場合はステップＥ２へ進み、学習対情報Ｐｋを受けとった場合はステップＥ３へ進む。

【０３１２】

ステップＥ２：　選択伝達情報の種別情報Ａの内容が、認識結果情報を表す記号「Ｒ」である場合には、伝達内容情報Ｂの内容を合成部２０６に渡し、種別情報Ａの内容が圧縮結果情報を表す記号「Ｃ」である場合には、伝達内容情報Ｂの内容を、伸長部２０７へ渡し、ステップＥ１へ戻る。

【０３１３】

ステップＥ３：　学習対情報Ｐｋの音声素片情報Ｄの内容Ｖｋを伸長部２０７へ送り、その処理結果である音声波形情報Ｗｋを得る。

【０３１４】

ステップＥ４：　合成規則記憶部２１１に新たなエントリＲｋを生成し、そのラベル情報Ａ、Ｆ０情報Ｂ、継続時間情報Ｃ、および音声素片波形情報Ｄに、学習対情報Ｐｋのラベル情報Ａの値、Ｆ０情報Ｂの値、継続時間情報Ｃの値、および音声波形情報Ｗｋをそれぞれ記録する。

【０３１５】

以上の処理によって、学習対情報として伝達された情報を用い、新しい合成規則が生成されることとなる。

【０３１６】
以上が本装置の第二の実施例の構成とその機能である。
【０３１７】
かくしてこのように学習機能を設けて構成された本システムによれば、認識結果のスコアが基準以上で、かつ、まだ学習されていない音素を含むことが分かった時に、学習対情報という形で、認識結果と圧縮結果の組を送信し、受信部で受けとって合成規則を学習し、次回以降はこの規則によって合成音で出力を行うことが可能になる。
【０３１８】
これにより、本装置を使用していく過程で次第に合成音声が入力音声に類似していくことになり、声の個性を保存し、誰が喋っているかわかるようになり、かつ、声が経時変部しても正しく伝送でき、かつ学習によって、従来は知的符号部によって通信できなかった音声が、知的符号部によって送信できるようになることで、徐々に通信効率が向上する音声伝送装置を実現することが出来るようになる。
【０３１９】
尚、第二の実施例は上述した例に限定されるものではない。例えば、上述の例では基本周波数と、継続時間長に閾値を設けることで、学習の実施あるいは非実施の判断を行う例を示したが、この閾値を調整することによって、圧縮の効果の度合と学習処理の効果の度合等を制御することが出来る。
【０３２０】
また、上述の例では、音声素片を処理の単位とする例を示したが、例えば、単語などを処理単位とする実現も可能である。
【０３２１】
また、上述の例では基本周波数と、継続時間長に基づいて、学習の実施あるいは非実施の判断を行う例を示したが、例えば、話者情報や時間情報などを追加し、本装置を通常の利用者でない他者が利用した場合や、合成規則が学習されてから長い時間が経過した際には、同じ音韻に対しても再度学習を行うようにして、利用者の変部や、経時変化に対しても影響を受けない装置を構成することも可能である。
【０３２２】
また、上述の例では、説明を簡単化するため、音声認識結果と音声圧縮結果が同期して出力される例を示したが、これを非同期として同様の効果を実現することも可能である。
【０３２３】
また、上述の例では、音声合成規則として、音声素片が学習される例を示したが、言語解析処理に使われる言語辞書情報を学習する様にしても良い。また、上述では、音声素片選択のキーとして、Ｆ０，継続時間長を利用する例を示したが、注目している音素の前後の音素のラベルを参照して、音素を選択する合成部のための学習機能を実現することも可能である。
【０３２４】
また、上述の例では、送信制御部２０４から送られた全ての学習対情報に基づく学習処理が短時間に正しく実施される学習部を持つ例を示したが、学習状況記憶部２０９の各エントリに学習の成否などを記録する欄を追加し、かつ、学習部２１０での学習処理の成否などを通信路を通じて受信部から送信部に通信し、学習が失敗した場合には学習状況記憶部２０９の対応するエントリを削除するように構成することも可能である。
【０３２５】
次に第一の実施例の構成要素に、検証機能を追加した構成を第三の実施例として説明する。
【０３２６】
［第三の実施例］
第三の実施例は、送信部において音声認識結果に基づき、一旦、音声合成処理を行い、入力音声との比較を行い、その結果に応じて、受信部に伝送する信号の種別を決定することにより、受信部における合成誤りの発生を未然に防ぎ、より信頼性の高い高能率な音声伝

送を実現するものである。

【０３２７】

図２０は、このような本発明の第三の実施例にかかる音声伝達装置の構成の概要を表しており、本装置が図１に示した第一の実施例の構成要素に、検証用合成部３１２、および検証部３１３を追加した構成を持っていることが示されている。

【０３２８】

本実施例の特徴は、入力音声を音声認識・音声合成技術により超低ビットレートで圧縮された復号音声と原音声の特徴量の比較を行い、その結果に応じて復号器に伝送するパラメータを切り替える点にある。

【０３２９】

本実施例によれば、音声認識・音声合成技術により超低ビットレートで圧縮された復号音声と原音声の特徴量の比較を行い、類似している場合には、認識結果情報を伝送パラメータとして復号器に伝送される。

【０３３０】

この場合、認識結果情報は超低ビットレートで表すことができるため、周波数資源の有効利用に大きく寄与することができる。

【０３３１】

また、音声認識・音声合成技術により超低ビットレートで圧縮された復号音声と原音声の特徴量が類似していない場合には、音声符号部技術により求められる符号化情報が伝送パラメータとして復号器に伝送される。この場合、認識結果情報を伝送するのではなく、原音声をより忠実に表す音声符号化技術にて求められた符号部パラメータを伝送することにより、品質の向上を図ることが可能となる。

【０３３２】

以下、図２１を用いて説明する。図２１は、本実施例の一構成を示すブロック図である。この構成においては、入力端子２１０１から音声信号が入力され、認識部３０２に与えられる。パターン認識部３０２では周波数分析がなされ、時間周波数で表される音声パターンが得られる。

【０３３３】

その音声パターンと予め作成してある標準パターンとの類似度を時間軸を非線形に伸縮させながら算出し、最も類似度の高い標準パターンの表す音素を認識結果情報として出力する。

【０３３４】

標準パターンの単位としては、音素の他に音節、単語などの種々のレベルが考えられる。

【０３３５】

認識部３０２で求められた認識結果情報は合成音声を生成するための圧縮部３０３に与えられる。圧縮部３０３では、認識結果情報から形態素解析、係り受け解析、韻律制御などを経て合成音声を生成し、類似度を算出するための類似度算出部３０４に与える。

【０３３６】

類似度算出部３０４では、圧縮部３０３で生成した合成音声と入力音声との類似度を算出する。ここでいう類似度とは、合成音声の特徴量と入力音声の特徴量の差を指す。

【０３３７】

特徴量として、例えば、ピッチパターン、サブバンドパワー、ＬＰＣスペクトル、ゼロ交叉比など音響パラメータの１つ以上の組み合わせを用いることができる。類似度算出部３０４で求めた合成音声と入力音声の類似度の情報を、当該判定合成音声と入力音声との類似度判定をするための判定部３０５に与え、当該判定部３０５では認識部３０２および圧縮部３０３による合成音声が入力音声に類似しているかを閾値等を用いて判定する。

【０３３８】

ここで類似していると判定された場合に、スイッチ２１０３は端子２１０２と接続し、スイッチ２１０６をオフ状態とし、認識部３０２で求められた認識結果情報をマルチプレクサ３０４ａに与える。このとき符号部３０３ａは動作しない。

【０３３９】
一方、判定部３０５において、認識部３０２および圧縮部３０３による合成音声が入力音声に類似していないと判定した場合には、スイッチ２１０３を端子２１０４に接続し、スイッチ２１０６を端子２１０５に接続して入力音声を符号部３０３ａに与えて符号部情報を求め、これをマルチプレクサ３０４ａに与える。
【０３４０】
マルチプレクサ３０４ａでは、認識結果情報または符号部情報の一方と判定部３０５より得られる判定情報とをビットストリームに変換し、出力端子２１０７より出力する。
【０３４１】
次に、本実施例の別の構成を図２２を用いて説明する。
【０３４２】
この構成は、音声認識・音声合成技術により超低ビットレートで圧縮された復号音声と原音声の特徴量の差と、音声符号部技術により圧縮された復号音声と原音声の特徴量の差を比較し、原音声との特徴量の差の小さい技術で求められた情報（認識結果情報または符号部情報）を伝送パラメータとする点に特徴がある。
【０３４３】
この構成の利点は、常に原音声の特徴量に近い情報（認識結果情報または符号部情報）が選択されるため、高品質で安定した復号音声を得ることが可能となる点にある。
【０３４４】
以下、図２２を用いて説明を行う。図２２において、図２１と同じ名称が付されている構成要素は同様の機能を有するものとし、ここでは説明を省略する。
【０３４５】
入力端子２２０１より音声信号が入力され、認識部３０２および符号部部３０３ｂに与えられる。認識部３０２では、認識処理を行い認識結果情報を求め合成部３１２ｂに与える。合成部３１２ｂでは合成音声を生成し類似度算出部３１３ｃに与える。類似度算出部３１３ｃでは、合成音声と入力音声の類似度を求め、その結果を比較部３１３ｅに与える。
【０３４６】
これと並行して、符号部３０３ｂでは入力信号をＣＥＬＰなどの技術を用いて符号部情報を求め、この符号部情報を復号部３０３ｃに与えて復号音声を求める。次に類似度算出部３１３ｄでは、復号部５３０３ｃから与えられる復号音声と入力音声との類似度を、類似度算出部３１３ｃと同様の手法により求め、その結果を比較部３１３ｅに与える。
【０３４７】
比較部３１３ｅでは類似算出部３１３ｃから選られる類似度と類似度算出部５　３１３ｄから得られる類似度の比較を行い、どちらの類似度が大きいかを判定し、その結果に応じてスイッチ２２０３を切り替える。
【０３４８】
類似度算出部３１３ｃからの類似度が大きい場合には、スイッチ２２０３は端子２２０２と接続して認識部３０２で求められた認識結果情報がマルチプレクサ３０４ｂに与えられる。
【０３４９】
類似度算出部３１３ｄからの類似度が大きい場合には、スイッチ２２０３は端子２２０４と接続して符号部３０３ｂで求められた符号化情報がマルチプレクサ３０４ｂに与えられる。マルチプレクサ３０４ｂでは、認識結果情報もしくは符号化情報のいずれか一方と、比較部３１３ｅで求められた比較結果が与えられ、ビットストリームに変換した後に出力端子２２０５より出力する。
【０３５０】
なお、本発明は上述した例に限定されるものではない。たとえば、本装置による通信は、双方向通信としても実現することが可能であるが、その際、伝達する認識結果情報の検証のために利用する検証用合成手段と、選択伝達情報として伝達されてきた認識結果情報に基づく音声合成処理を行う音声合成手段とを一つの合成装置によって実現することにより

、装置の小型化、簡素化を図ることも可能である。すなわち、双方向通信にした場合の資源の共有ことにより、装置の小型化、簡素化を図ることが可能である。

【０３５１】

以上、第三の実施例で説明した本発明にかかる音声伝達装置は、送信部において音声認識結果に基づき一旦音声合成処理を行い、入力音声との比較を行い、その結果に応じて、受信部に伝送する信号の種別を決定するようにしたことにより、受信部における合成誤りの発生を未然に防ぎ、より信頼性の高い高能率な音声伝送を実現することができるものである。

【０３５２】

次に第四の実施例を説明する。

【０３５３】

［第四の実施例］

第四の実施例では、送信部において音声認識結果に基づく高精度な音声合成処理と、受信部において行われる音声合成処理の双方の出力を比較し、その結果に応じて、送信部側から受信部に伝送する信号の種別を決定することで、受信部における合成誤りの発生を未然に防ぎ、より信頼性の高い高能率な音声伝送を実現する。すなわち、認識結果を送信した場合に受信部で正しく合成できるかどうかを認識結果の送信前に送信装置側で判定した上で、正しく合成できそうな場合は認識結果を送るようにし、そうでない場合には圧縮結果を送るようにすることにより、受信装置側での合成誤りを抑制する。

【０３５４】

これによって、受信部に於ける合成誤りの発生を未然に防ぎ、より信頼性の高い高能率な音声伝送を実現するものである。

【０３５５】

以下、図面を参照して本発明の第四の実施例に係る音声伝達装置につき説明する。図２３は、本発明の第四の実施例にかかる音声伝達装置の構成の概要を表しており、本装置が図１に示した第一の実施例の構成要素に、検証用合成部４１２、および検証部４１３および高品位合成部４１４を追加した構成を持っていることが示されている。すなわち、第一の実施例の構成要素に、高性能合成器を用いての検証機能を付加した構成である。

【０３５６】

ここで、入力部４０１、認識部４０２、圧縮部４０３、受信制御部４０５、合成部４０６、伸長部４０７、および出力部４０８は、第一の実施例の入力部４０１、認識部１０２、圧縮部１０３、受信制御部１０４５、合成部１０６、伸長部１０７、出力部１０８と同様の機能を持つ。

【０３５７】

検証用合成部４１２は、認識部４０２の出力する認識結果情報を受けとり、合成部４０６と同様の音声合成処理の少なくとも一部の処理を行い、形態素解析結果、あるいは係り受け解析結果、あるいは音素ラベル系列、あるいは韻律制御指令、あるいは合成音声出力波形などといった音声合成処理のための中間データあるいは音声合成処理結果の少なくとも一つを検証用出力情報として出力し、検証部４１３に渡すようにしている。

【０３５８】

高品位合成部４１４は、認識部４０２の出力する認識結果情報を受けとり、大規模な言語解析規則や、韻律決定規則や、音声素片の情報を用いることなどによって、合成部４０６より高精度な音声合成処理の少なくとも一部の処理を行い、形態素解析結果、あるいは係り受け解析結果、あるいは音素ラベル系列、あるいは韻律制御指令、あるいは合成音声出力波形などといった音声合成処理所ための中間データあるいは音声合成処理結果の少なくとも一つを基準出力情報として出力し、検証部４１３に渡すようにしている。

【０３５９】

検証部４１３は、検証用合成部４１２から得られる検証用出力情報と、高品位合成部４１４から得られる基準出力情報を比較し、両者にあらかじめ定めた基準以上の相違がある場合には、送信制御部４０４に対応する音声入力は圧縮結果情報を選択伝達情報として送信

させるよう制御し、そうでない場合には、送信制御部４０４で行われる従来の基準にしたがって、対応する音声入力を、圧縮結果情報あるいは認識結果情報を選択伝達情報として送信させるよう構成する。

【０３６０】

つづいて具体的に処理内容を説明する。

【０３６１】

いま、入力部４０１から「寝台車頼む」という音声入力がなされたとする。

【０３６２】

この音声信号が認識部４０２において認識処理され、認識結果情報として例えば「／し／ん／だ／い／しゃ／た／の／む／」という音素ラベル列が得られ、検証用合成部４１２および高品位合成部４１４に渡されたとする。

【０３６３】

この認識結果情報に対し、検証用合成部４１２での処理によって、「読み：　（しんだ）、アクセント：（０型）、品詞：動詞／読み：（いしゃ）、アクセント：（０型）、品詞：名詞／読み：（たのむ）、アクセント：（２型）、品詞：動詞」という検証用出力情報が得られ、検証部４１３に渡されたものとする。　（これは入力が「死んだ医者頼む」と誤って解析されたことに対応する。）

一方、同じ認識結果情報に対して、高品位合成部４１４での処理によっては、「読み：（寝台）、アクセント：（３型）、品詞：名詞／読み：（たのむ）、アクセント：（２型）、品詞：動詞」という基準出力情報が得られ、検証部４１３に渡されたものとする。（これは入力が「寝台車頼む」と正しく解析されたことに対応する。）

検証部４１３では、上述の検証用出力情報と、基準出力情報を受けとり比較を行うが、これらの間に形態素解析による品詞の割り振りに相違があり、その結果アクセント指令にも相違が生じていることが判明する。

【０３６４】

仮に、この認識結果情報を選択伝達情報として伝送すると合成部４０６において誤った合成がなされ、フレージングおよびアクセントに関して問題のある出力がなされることを事前に検知したので、送信制御部４０４を制御して、入力音声に対応する圧縮結果情報を含む選択伝達情報が送信される。

【０３６５】

これによって、合成処理の誤りに起因する音声の伝達誤りを起こさない音声伝達装置を構成することが出来る。

【０３６６】

このように、第四の実施例では、送信部において音声認識結果に基づき、高精度な音声合成処理を行い、また、受信部において音声合成処理を行い、これら送信部および受信部での音声処理出力を比較し、その結果に応じて、受信部に伝送する信号の種別を決定するようにした。これにより、受信部における合成誤りの発生を未然に防ぎ、より信頼性の高い高能率な音声伝送を実現することができるようになる。

【０３６７】

すなわち、認識結果を送信した場合に受信部で正しく合成できるかどうかを認識結果の送信前に送信装置側で判定した上で、正しく合成できそうな場合は認識結果を送るようにし、そうでない場合には圧縮結果を送るようにしたことにより、受信部に於ける合成誤りの発生を未然に防ぎ、より信頼性の高い高能率な音声伝送を実現することができるようになる。

【０３６８】

次に、話者照合手段によって、発声者が誰であるかを特定し、話者情報を通信路を介して送し、合成手段において音声に変換する場合に、話者の声に近い音声を出力することができるようにした実施例を第五の実施例として説明する。

【０３６９】

［第五の実施例］

以下、図面を参照して本発明の第五の実施例に係る話者照合機能追加の音声伝達装置につき説明する。

【０３７０】

図２４は、本発明の第五の実施例にかかる音声伝達装置の構成の概要を表しており、本装置が図１に示した第一の実施例の構成要素に、話者を認証する話者認証部５１５を追加し、また、受信側の合成部５０６にはこの話者認証情報を元に話者の声に近い音声を生成する機能を持たせている。

【０３７１】

ここで、送信部側に設けた話者認証部５１５では、予め作成してある話者ごとの特徴を示す話者特徴辞書を有しており、これを用いて発声者が誰であるかを特定し、話者情報を通信路を介して送る機能を有している。すなわち、話者認証部５１５には入力された音声データを周波数分析し、時間周波数パターンで表される音声パターンが得、その音声パターンと、予め作成してある話者ごとの特徴を示す話者特徴辞書との照合を行い、各話者ごとにどのくらい似ているかを表すスコアを計算し、最も高いスコアを示す話者を照合結果（話者認証情報）として出力する機能を持たせてある。また、合成部５０６は話者の素片辞書を有しており、送られてきた上記話者情報に対応する話者の素片辞書を選択し、その素片を用いて音声合成処理を行う。

【０３７２】

これにより、受信部では送信部から選択伝達情報として送られて来た認識結果情報を合成部５０６により音声に変換する場合に、話者の声に近い音声を生成することも可能となる。

【０３７３】

つづいて、具体的に処理内容を説明する。

【０３７４】

まず、話者認証部５１５に入力された音声データは、認識部５０２と同様に、周波数分析がなされ、時間周波数パターンで表される音声パターンが得られる。

【０３７５】

話者認証部５１５ではその音声パターンと、予め作成してある話者ごとの特徴を示す話者特徴辞書との照合を行い、各話者ごとにどのくらい似ているかを表すスコアを計算する。そして、最も高いスコアを示す話者が照合結果として出力される。

【０３７６】

こうして得られた話者情報は通信路を介して話者認証部５１５から、受信制御部５０５へと送られる。

【０３７７】

ここで、合成部５０６による音声合成処理によって出力する旨の選択がなされた場合、合成部５０６では送られてきた上記話者情報に対応する話者の素片辞書を選択し、その素片を用いて音声合成処理を行う。これににより、聞き手は話し手の声に近い音声を聞くことができる。

【０３７８】

ここで、話者照合に用いる話者特徴辞書および音声合成で用いる音声素片辞書は、対応づいている必要があるが、同一人物の声である必要はない。

【０３７９】

また、上記辞書は発声者の声で作成しておく必要もない。ただし、用意しておく話者の数を増やすことにより、より発声者に近い音声を合成することができる。

【０３８０】

例えば、話者特徴辞書として５人の特徴を登録しておいた場合に、未知話者が発声した音声から５人のうち、どの話者に近いかを求めて、最も近い人の音声で合成音を生成するよりも、１００人登録しておき、その中から最も近い人を選択した方が、発声者により近い合成音になる可能性が高い。

【０３８１】

また、話者特徴辞書はＡ～Ｅの５人の話者から作成しておき、音声合成素片辞書は話者Ｆ～Ｊの５人で作成しておき、その間で最も近い話者の対応表を予め作成しておく。これにより、話者Ａ～Ｅのいずれかが選択された場合に、対応する人の素片辞書を用いて合成音を生成することで、発声者に近い音声を再現できるようになる。

【０３８２】

以上の第五の実施例は、話者照合手段によって発声者が誰であるかを特定し、話者情報を通信路を介して伝送して合成手段に渡すようにしたので、合成手段において音声に変換する場合に、話者の声に近い音声を出力することができるようになる。

【０３８３】

以上、種々の実施例を説明したが、いずれも上述の実施例では、装置として本発明を実現する場合のみを示した。しかし、上述の具体例の中で示した処理手順あるいはフローチャートを、プログラムとして記述し、実装し、汎用の計算機システムに読み込んで実行することによっても、同様の機能と効果を得ることが可能である。

【０３８４】

すなわち、この場合、図２５の汎用コンピュータの構成の例に示したように、入力インタフェース６０１、出力インタフェース６０２、ＣＰＵ６０３、メモリ６０４、大容量記憶装置６０５、および通信インタフェース６０７からなる汎用コンピュータを、通信路６０７によって接続し、ＣＰＵによるソフトウエア制御により、上述のごとき動作を実現することが出来る。

【０３８５】

すなわち、上述の実施例に記載した手法は、コンピュータに実行させることの出来るプログラムとして、磁気ディスク（フロッピディスク、ハードディスクなど）、光ディスク（ＣＤ－ＲＯＭ、ＤＶＤ、ＭＯなど）、半導体メモリなどの記録媒体を用いてコンピュータにプログラムを読み込み、ＣＰＵ６０３で実行させれば、本発明の音声伝送装置を実現することが出来ることになる。

【０３８６】

【発明の効果】

以上説明したように、本発明によれば、従来の符号化による圧縮通信では、伝達できる音声の品質と圧縮度の間にトレードオフがあるため、可能な圧縮比率に限界があるという問題点を解決し、より効率の高い音声伝送を実現する。

【０３８７】

また、従来の知的符号化による通信では避けることの出来なかった、音声認識処理での誤認識に起因する伝送内容の誤りを、避けることが出来る音声伝送を実現する。

【０３８８】

また、従来の知的符号化による通信では避けることの出来なかった、音声合成処理での誤生成に起因する伝送内容の誤りを、避けることが出来る音声伝送を実現する。

【０３８９】

また、以上によって、従来の符号化による圧縮通信より高い効率の伝送を、伝送内容に誤りを起こさずに実現するものである。

【０３９０】

また、高い通信効率を達成しつつ、本装置を利用する利用者の声の個人性を保存し誰が喋っているかかわかる音声伝達装置を実現する。

【０３９１】

また、声が経時変化しても随時学習によって追従する音声伝達を実現し、また、随時学習を行うことによって、利用するほどに段々通信効率が上がる音声伝達を実現するものである。また、検証機能によって信頼性の高い知的符号化を実現する等の実用上多大な効果が奏せられる。

【図面の簡単な説明】

【図１】本発明の第一の実施例にかかる音声伝達装置の構成の概要を説明するための図。

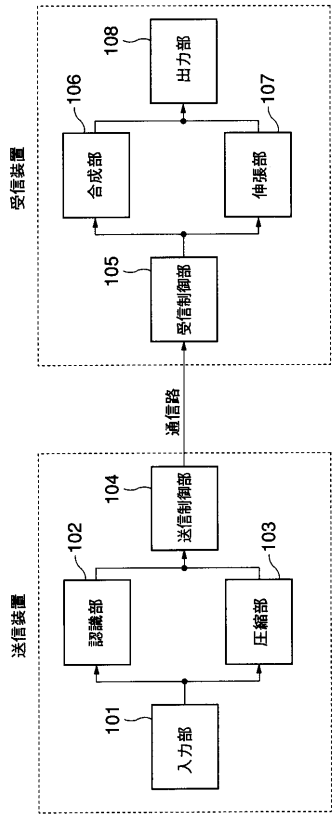
【図２】ＣＥＬＰ符号化のエンコーダ部を表すブロック図。

【図３】ＣＥＬＰ符号化のデコーダ部を表すブロック図。
【図４】認識結果情報の内容の例を示す図。
【図５】圧縮結果情報の内容の例を示す図。
【図６】選択伝達情報の内容の例を示す図。
【図７】合成音部の構成の例を示す図。
【図８】合成音声出力情報の内容の例を示す図。
【図９】伸長音声出力情報の内容の例を示す図。
【図１０】送信制御部１０４の内部構成の例を示す図。
【図１１】処理手順Ａを説明するフローチャート。
【図１２】認識結果情報および圧縮結果情報の時間的関係の例を示す図。
【図１３】本発明の第二実施例にかかる音声伝達装置の構成の概要を示す図。
【図１４】認識部２０２の出力する認識結果情報の例を示す図。
【図１５】圧縮部２０３の出力する圧縮結果情報の例を示す図。
【図１６】学習状況記憶部２０９の内容の例を示す図。
【図１７】合成規則記憶部２１１の内容の例を示す図。
【図１８】合成部２０６および学習部２１０および合成規則記憶部２１１の内部構成の例を示す図。
【図１９】学習対情報の例を示す図。
【図２０】本発明の第三実施例にかかる音声伝達装置の構成の概要を示す図。
【図２１】検証機能付きの送信部の一つの構成を示す図。
【図２２】検証機能付きの送信部の別の構成を示す図。
【図２３】本発明の第四実施例にかかる音声伝達装置の構成の概要を示す図。
【図２４】本発明の第五実施例にかかる音声伝達装置の構成の概要を示す図。
【図２５】汎用コンピュータの構成の例を示す図。
【符号の説明】
１０１…入力部
１０２…認識部
１０３…圧縮部
１０３ａ…スペクトル包絡符号帳探索部
１０３ｂ…適応符号帳探索部
１０３ｃ…雑音符号帳探索部
１０３ｄ…ゲイン符号帳探索部
１０３ｅ…多重化部
１０４…送信制御部
１０４ａ…認識情報キュー
１０４ｂ…圧縮情報キュー
１０４ｃ…送信部
１０４ｄ…制御処理部
１０５…受信制御部
１０６…合成部
１０６ａ…言語処理部
１０６ｂ…韻律処理部
１０６ｃ…音声波形生成部
１０６ｄ…言語辞書記憶部
１０６ｅ…韻律規則記憶部
１０６ｆ…音声単位情報記憶部
１０７…伸長部
１０７ａ…適応音源復号部
１０７ｂ…雑音音源復号部
１０７ｃ…ゲイン音源復号部

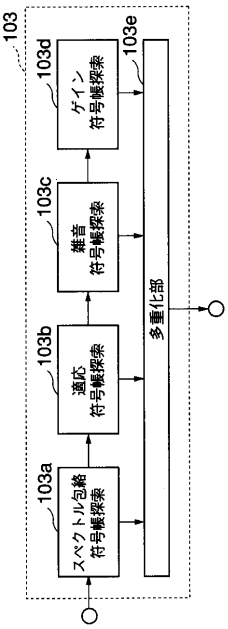
１０７ｄ…スペクトル包絡復号部
１０７ｅ…音源再生部
１０７ｆ…合成フィルタ
１０７ｑ…ポストフィルタ
１０７ｈ…逆多重化部
１０８…出力部
２０１…入力部
２０２…認識部
２０３…圧縮部
２０４…送信制御部
２０５…受信制御部
２０６…合成部
２０６ａ…言語処理部
２０６ｂ…韻律処理部
２０６ｃ…音声波形生成部
２０７…伸長部
２０８…出力部
２０９…学習状況記憶部
２１０…学習部
２１０ａ…音声分析部
２１０ｂ…韻律規則学習部
２１０ｃ…音声単位情報学習部
２１１…合成規則記憶部
２１１ａ…言語辞書記憶部
２１１ｂ…韻律規則記憶部
２１１ｃ…音声単位情報記憶部
３０１…入力部
３０２…認識部
３０３…圧縮部
３０３ａ…符号化部
３０３ｂ…符号化部
３０３ｃ…復号部
３０４…送信制御部
３０４ａ，３０４ｂ…マルチプレクサ
３０５…受信制御部
３０６…合成部
３０７…伸長部
３０８…出力部
３１２…検証用合成部
３１３…検証部
３１３ａ…類似度算出部
３１３ｂ…判定定部
３１３ｃ…類似度算出部
３１３ｄ…類似度算出部
３１３ｅ…比較部
４０１…入力部
４０２…認識部
０３…圧縮部
４０４…送信制御部
４０５…受信制御部

４０６…合成部
４０７…伸長部
４０８…出力部
４１２…検証用合成部
４１３…検証部
４１４…高品位合成部
５０１…入力音
５０２…認言１部
５０３…圧縮部
５０４…送信制御部
５０５…受信制御部
５０６…合成部
５０７…伸長部
５０８…出力部
５１５…話者照合部
６０１ａ，６０１ｂ…入力インタフェース
６０２ａ，６０２ｂ…出力インタフェース
６０３ａ，６０３ｂ…ＣＰＵ
６０４ａ，６０４ｂ…メモリ
６０５ａ，６０５ｂ…大容量記憶装置
６０６ａ，６０６ｂ…通信インタフェース
６０７…通信路
２１０１…入力端子
２１０２…端子
２１０３…スイッチ
２１０４，２１０５，２２０２，２２０４…端子
２１０６，２２０３…スイッチ
２１０７，２２０５…出力端子
２２０１…入力端子。

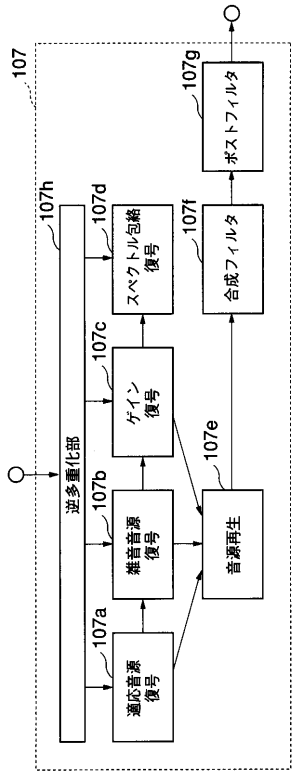
【図１】



【図２】



【図３】



【図４】

認識結果情報の内容の例

ID	始端情報A	終端情報B	認識内容情報C	スコア情報D
:	:	:	:	:
I1	T1s	T1e	<わたしは>	S1
I2	T2s	T2e	<さとう>	S2
I3	T3s	T2e	<いちろう>	S3
I4	T4s	T3e	<です>	S4
:	:	:	:	:

【図５】

圧縮結果情報の内容の例

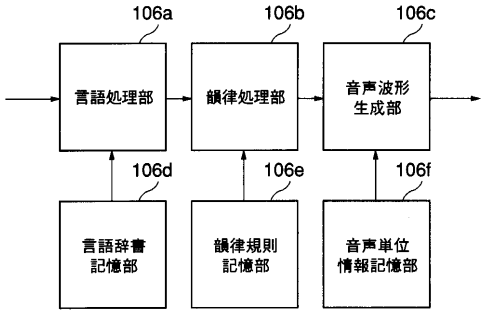
ID	始端情報A	終端情報B	圧縮内容情報C
:	:	:	:
J1	U1s	U1e	(..わたしわ..)
J2	U2s	U2e	(..さとう..ひち)
J3	U3s	U3e	(ろう..です..)
:	:	:	:
J1'	T1s	U1e	(わたしわ..)
J1''	T1e	U1e	(..)
:	:	:	:
J2'	T2s	U2e	(さとう..ひち)
J2''	T2e	U2e	(..ひち)
:	:	:	:
J3'	T4s	U3e	(です..)
J3''	T4e	U3e	(..)
:	:	:	:

【図６】

選択伝達情報の内容の例

ID	種別情報A	伝達内容情報B
:	:	:
K1	C	[J1 U1s T1s (..)]
K2	R	[I1 T1s T1e <わたしは> S1]
K3	C	[J1'' T1e U1e (..)]
K4	C	[J2 U2s T2s (..)]
K5	R	[I2 T2s T2e <さとう> S2]
K6	C	[J2'' T2e U2e (..ひち)]
K7	C	[J3 U3s T4s (ろう..)]
K8	R	[I4 T4s T4e <です> aaS3]
K9	C	[J3'' T4e U3e (..)]
:	:	:

【図７】



【図８】

合成音声出力情報の内容の例

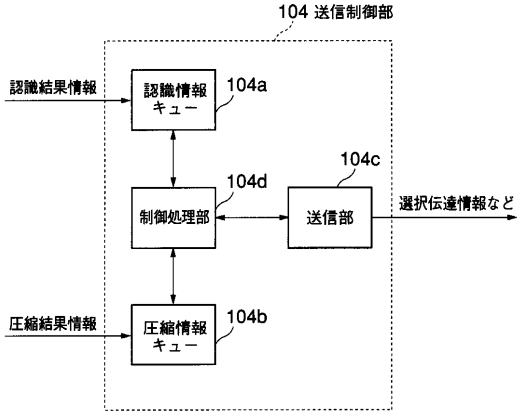
ID	始端情報A	終端情報B	合成音声波形情報C
:	:	:	:
M1	T1s	T1e	{わたしは}
M2	T2s	T2e	{さとう}
M3	T4s	T3e	{です}
:	:	:	:

【図９】

伸張音声出力情報の内容の例

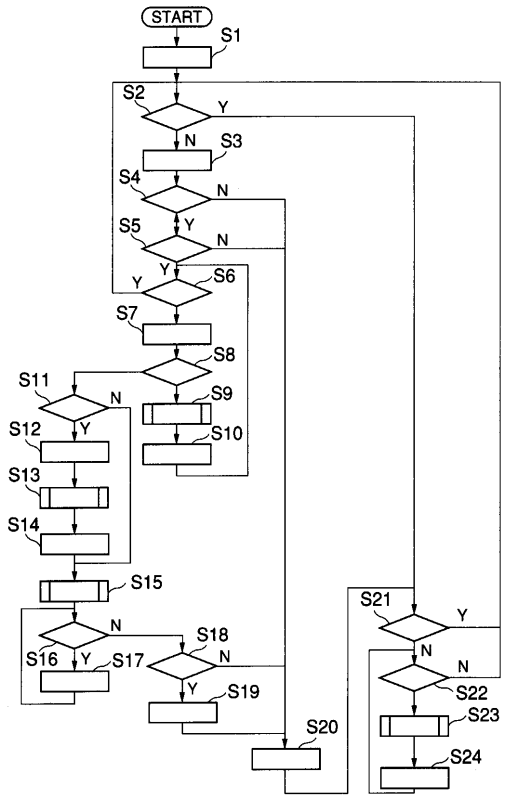
ＩＤ	始端情報A	終端情報B	伸張音声波形情報C
：	：	：	：
N1	U1s	T1s	{..}
N2	T1e	U1e	{..}
N3	U2s	T2s	{..}
N4	T2e	U2e	{..ひち}
N5	U3s	T4s	{ろう..}
N6	T4e	U3e	{..}
：	：	：	：

【図１０】

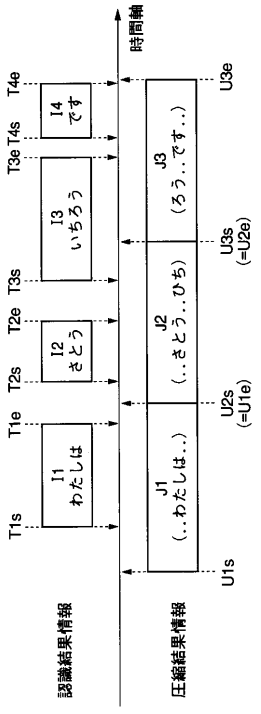


【図１１】

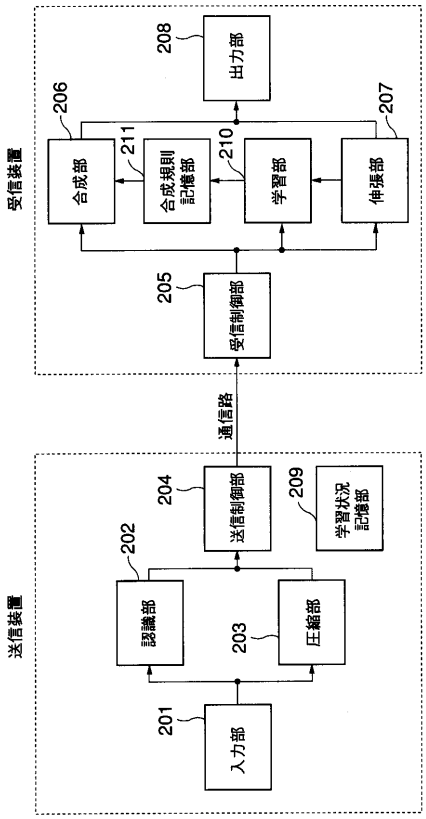
処理手順Aを説明するフローチャート



【図１２】



【図１３】



【図１４】

認識部202の出力する認識結果情報の例

ID情報	始点情報A	終点情報B	ラベル情報C	スコア情報D	FO情報E
：	：	：	：	：	：
I1'	T1s'	T1e'	/こ/	S1'	F1
I2'	T2s'	T2e'	/ん/	S2'	F2
I3'	T3s'	T3e'	/に/	S3'	F3
I4'	T4s'	T4e'	/ち/	S4'	F4
I5'	T5s'	T5e'	/わ/	S5'	F5
：	：	：	：	：	：

【図１５】

圧縮部203の出力する圧縮結果情報の例

ID	始端情報A	終端情報B	圧縮内容情報C
：	：	：	：
J1'	T1s'	T1e'	(こ)
J2'	T2s'	T2e'	(ん)
J3'	T3s'	T3e'	(に)
J4'	T4s'	T4e'	(ち)
J5'	T5s'	T5e'	(わ)
：	：	：	：

【図１６】

学習状況記憶部209の内容の例

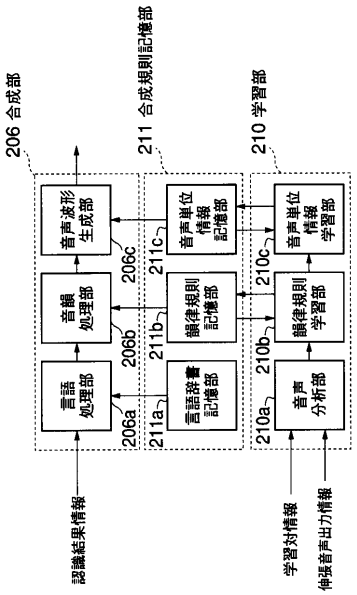
ID情報	ラベル情報A	FO情報B	継続時間情報C	音声素片情報D
：	：	：	：	：
X1	/こ/	F1	D1	(こ)
X2	/ん/	F2	D2	(ん)
X3	/に/	F3	D3	(に)
X4	/ち/	F4	D4	(ち)
X5	/わ/	F5	D5	(わ)
：	：	：	：	：

【図１７】

合成規則記憶部211の内容の例

ID情報	ラベル情報A	FO情報B	継続時間情報C	音声素片波形情報D
：	：	：	：	：
Y1	/こ/	F1	D1	{こ}
Y2	/ん/	F2	D2	{ん}
Y3	/に/	F3	D3	{に}
Y4	/ち/	F4	D4	{ち}
Y5	/わ/	F5	D5	{わ}
：	：	：	：	：

【図１８】

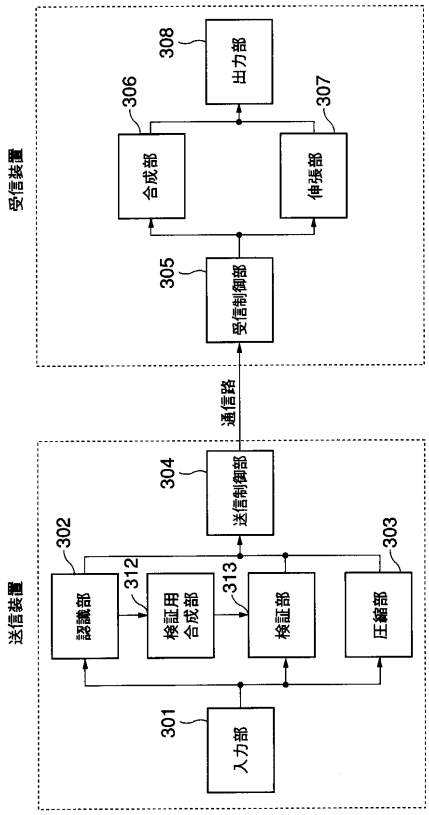


【図１９】

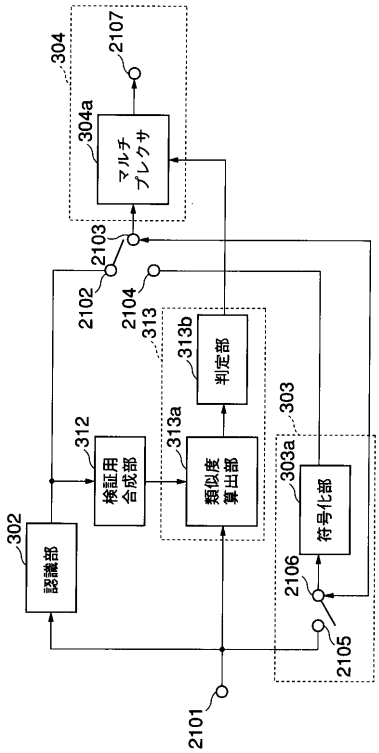
学習対情報の例

ID情報	ラベル情報A	FO情報B	継続時間情報C	音声素片情報D
：	：	：	：	：
Z1	/こ/	F1	D1	(こ)
Z2	/ん/	F2	D2	(ん)
Z3	/に/	F3	D3	(に)
Z4	/ち/	F4	D4	(ち)
Z5	/わ/	F5	D5	(わ)
：	：	：	：	：

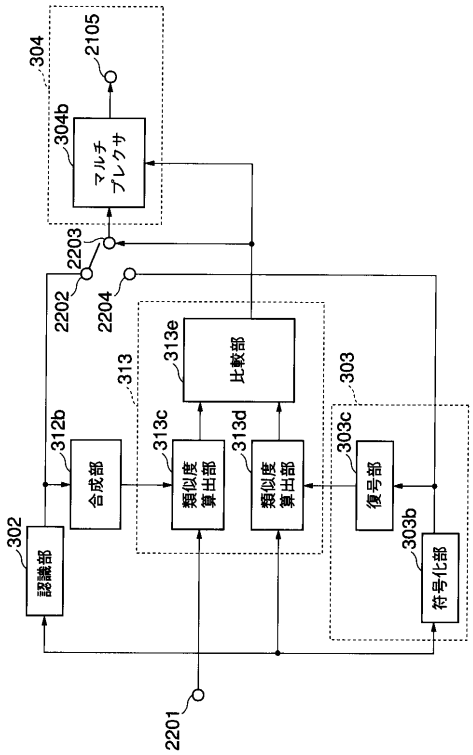
【図２０】



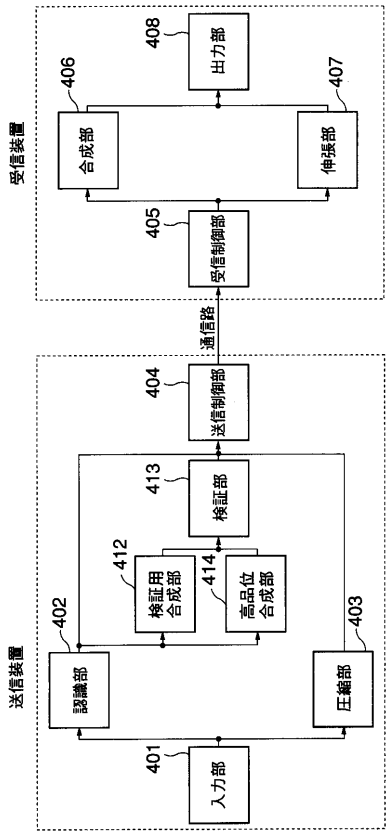
【図２１】



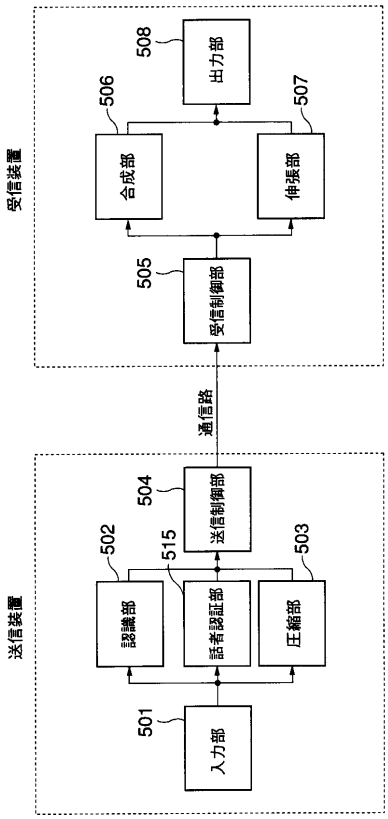
【図２２】



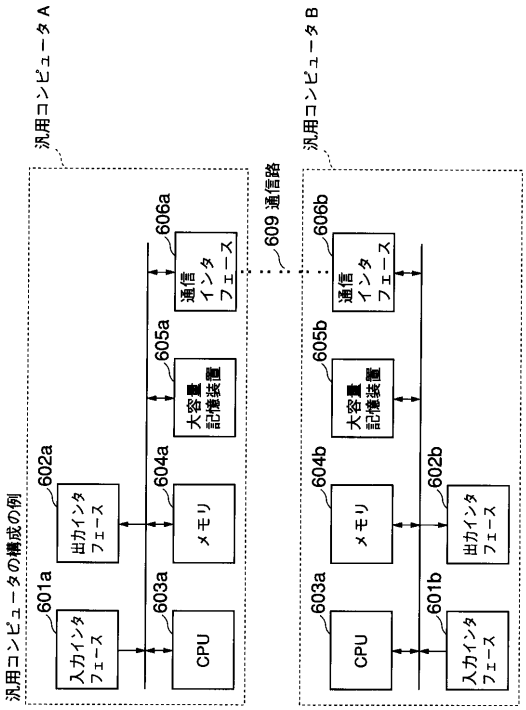
【図２３】



【図２４】



【図２５】


汎用コンピュータの構成の例
汎用コンピュータＡ
汎用コンピュータＢ
609 通信路
601a
入力インタフェース
602a
出力インタフェース
603a
CPU
604a
メモリ
605a
大容量記憶装置
606a
通信インタフェース
601b
入力インタフェース
602b
出力インタフェース
603a
CPU
604b
メモリ
605b
大容量記憶装置
606b
通信インタフェース

フロントページの続き

(51)Int.Cl.[7]		FI		
Ｇ１０Ｌ	15/14	Ｇ１０Ｌ	9/18	３０１Ｂ
Ｇ１０Ｌ	17/00	Ｇ１０Ｌ	3/00	５３５Ｚ
Ｇ１０Ｌ	19/12	Ｇ１０Ｌ	3/00	５４５Ａ
Ｈ０３Ｍ	7/30	Ｇ１０Ｌ	3/00	５５１Ａ

(74)代理人　100070437
弁理士　河井　将次
(72)発明者　知野　哲朗
兵庫県神戸市東灘区本山南町８丁目６番２６号　株式会社東芝関西研究所内
(72)発明者　押切　正浩
兵庫県神戸市東灘区本山南町８丁目６番２６号　株式会社東芝関西研究所内
(72)発明者　金澤　博史
兵庫県神戸市東灘区本山南町８丁目６番２６号　株式会社東芝関西研究所内
(72)発明者　新居　孝章
兵庫県神戸市東灘区本山南町８丁目６番２６号　株式会社東芝関西研究所内
(72)発明者　笹島　宗彦
兵庫県神戸市東灘区本山南町８丁目６番２６号　株式会社東芝関西研究所内

審査官　渡邊　聡

(56)参考文献　特開平１０－２８５２７５（ＪＰ，Ａ）
特開平０４－０１０７３４（ＪＰ，Ａ）
特開昭６１－０４３７９６（ＪＰ，Ａ）
特開昭５７－０８８５００（ＪＰ，Ａ）
特表平０７－５０５２３７（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G10L 13/00
G10L 13/08
G10L 15/00
G10L 15/10
G10L 15/14
G10L 17/00
G10L 19/00
G10L 19/12
H03M 7/30