

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3846300号  
(P3846300)

(45) 発行日 平成18年11月15日(2006.11.15)

(24) 登録日 平成18年9月1日(2006.9.1)

(51) Int. Cl. F I  
**G 1 0 L 13/06 (2006.01)** G 1 0 L 13/06 2 3 0 Z  
**G 1 0 L 15/04 (2006.01)** G 1 0 L 15/04 2 0 0

請求項の数 5 (全 35 頁)

(21) 出願番号	特願2001-382100 (P2001-382100)	(73) 特許権者	000002945
(22) 出願日	平成13年12月14日 (2001.12.14)		オムロン株式会社
(65) 公開番号	特開2003-186489 (P2003-186489A)		京都市下京区塩小路通堀川東入南不動堂町
(43) 公開日	平成15年7月4日 (2003.7.4)		801番地
審査請求日	平成16年5月19日 (2004.5.19)	(74) 代理人	100080322
			弁理士 牛久 健司
		(74) 代理人	100104651
			弁理士 井上 正
		(74) 代理人	100114786
			弁理士 高城 貞晶
		(72) 発明者	平山 裕司
			京都市下京区塩小路通堀川東入南不動堂町
			801番地 オムロン株式会社内

最終頁に続く

(54) 【発明の名称】 録音原稿作成装置および方法

(57) 【特許請求の範囲】

【請求項1】

複数の文字列を含む元原稿を設定する手段、

元原稿に含まれる文字列を構成するすべての音声単位を抽出し、かつそれらの音声単位について、その元原稿における出現回数を検出する元原稿分析手段、および

上記元原稿分析手段によって抽出された音声単位の出現回数の検出結果に基づいて、出現回数の少ない音声単位を含む文字列から順に、上記すべての音声単位を網羅するまで、元原稿から文字列を選択して録音原稿を作成する第1の文字列選択手段、  
を備えた録音原稿作成装置。

【請求項2】

既存の音声情報データベースに含まれるすべての第1の音声単位を抽出する音声情報データベース分析手段、

追加元原稿に含まれる文字列を構成するすべての第2の音声単位を抽出する追加元原稿分析手段、

第2の音声単位のうち、第1の音声単位に含まれていない音声単位を検出する比較手段、および

上記比較手段によって検出された音声単位を含む文字列を追加元原稿から選択して追加録音原稿を作成する第2の文字列選択手段、

をさらに備えた請求項1に記載の録音原稿作成装置。

【請求項3】

10

20

元原稿分析手段および第 1 の文字列選択手段を備えた録音原稿作成装置の制御方法であって、

上記元原稿分析手段は、元原稿に含まれる文字列を構成するすべての音声単位を抽出し、かつそれらの音声単位について、その元原稿における出現回数を検出し、

上記第 1 の文字列選択手段は、出現回数の少ない音声単位を含む文字列から順に、上記の抽出したすべての音声単位を網羅するまで、元原稿から文字列を選択して録音原稿を作成する、

録音原稿作成装置の制御方法。

【請求項 4】

上記録音原稿作成装置は音声情報データベース分析手段、追加元原稿分析手段、比較手段および第 2 の文字列選択手段をさらに備えており、

上記音声情報データベース分析手段は、既存の音声情報データベースに含まれるすべての第 1 の音声単位を抽出し、

上記追加元原稿分析手段は、追加元原稿に含まれる文字列を構成するすべての第 2 の音声単位を抽出し、

上記比較手段は、第 2 の音声単位のうち、第 1 の音声単位に含まれていない音声単位を検出し、

上記第 2 の文字列選択手段は、検出された音声単位を含む文字列を追加元原稿から選択して追加録音原稿を作成する、

請求項 3 に記載の録音原稿作成装置の制御方法。

【請求項 5】

元原稿分析手段および文字列選択手段を備えた録音原稿作成装置を制御するプログラムであって、

上記元原稿分析手段が、元原稿に含まれる文字列を構成するすべての音声単位を抽出し、かつそれらの音声単位について、その元原稿における出現回数を検出し、

上記文字列選択手段が、出現回数の少ない音声単位を含む文字列から順に、上記の抽出したすべての音声単位を網羅するまで、元原稿から文字列を選択して録音原稿を作成するように制御するプログラム。

【発明の詳細な説明】

【0001】

【技術分野】

この発明は、音声対話の分野における、主に音声合成のための、音声情報データベース作成装置、さらにこのシステムの一部として位置づけられる録音原稿作成装置、録音管理装置、ラベリング装置、および方法に関する。

【0002】

従来、音声情報データベースの作成は、言語処理、録音、音声波形のラベリングなどの各分野に関する専門知識や技能を有した人間が手間と時間をかけて行っていたため、音声情報データベースの作成は、各分野の専門家を有する、または集めることのできる特定の企業や機関だけが行うことができた。また、必要な技能と時間の観点からも、音声情報データベース作成に必要なコストは膨大なものであった。

【0003】

この問題点はさらに、所望の音声で音声情報データベースを作成しさえすれば、あたかもその人が発話しているかのような自然な声質の合成音声を得られるという、波形接続型音声合成の特長が十分に活かされず、音声合成が広く普及しない要因にもなっていた。

【0004】

【発明の開示】

この発明は、音声合成により作成したいすべての文字列を含む元原稿から、音声情報データベースに音声情報を格納すべき最小限の文字列を持つ録音原稿を作成する装置および方法を提供するものである。

【0005】

10

20

30

40

50

この発明はまた、上記録音原稿を話者が声を出して発音することにより得られる音声波形に基づいて作成された音声情報データベースが既に存在するときに、さらに追加すべき文字列を含む追加原稿について、最小限の文字列を持つ追加録音原稿を作成する装置および方法を提供するものである。

【0006】

この発明はまた、録音原稿を話者が声を出して読むことを支援する録音管理装置および方法を提供するものである。

【0007】

この発明はさらに、録音により得られる音声波形とそれに対応する文字列とを用いて、音声波形についてのラベル情報（音声単位のラベルとその時間情報とからなる）を作成するとともに、作成したラベル情報の信頼性を高めることのできるラベリング装置および方法を提供するものである。

10

【0008】

この発明は最終的には、専門的知識をもたない者であっても、比較的容易に音声情報データベースを作成することができるシステムを提供するものである。

【0009】

この発明による録音原稿作成装置は、複数の文字列を含む元原稿を設定する手段、元原稿に含まれる文字列を構成するすべての音声単位を抽出する元原稿分析手段、および上記元原稿分析手段によって抽出されたすべての音声単位を含むように元原稿から文字列を選択して録音原稿を作成する第1の文字列選択手段を備えているものである。

20

【0010】

上記元原稿設定手段とは、手入力される元原稿を受付けるものでもよいし、FD等の記録媒体に格納されたものを読取るリーダでもよい。文字列とは、単語、句、節、文を含む概念である。いずれにしても元原稿には複数の文字列がある。第1の文字列選択手段は元原稿から選択する文字列の数が最小に（または、できるだけ少なくなるように）選択するものである。

【0011】

元原稿から選択して録音原稿に加えるべき文字列を最小にするためのこの発明の実施態様においては、上記元原稿分析手段は、元原稿に含まれる文字列を構成するすべての音声単位について、その元原稿における出現回数を検出するものであり、上記文字列選択手段は、出現回数の少ない音声単位を含む文字列から順に、上記すべての音声単位を網羅するまで、文字列を選択するものである。

30

【0012】

このようにして、この発明によると、音声情報データベースを作成するにあたって話者が声を出して読むべき録音原稿を元原稿から、操作者が言語処理の専門知識を有していなくても、作成することができる。しかも、録音原稿には（できるだけ）最小限の文字列のみが含まれることとなる。

【0013】

一実施態様では、上記文字列選択手段は、作成すべき音声情報データベースに関する所与の仕様を満たす条件の下で文字列を選択するものである。

40

【0014】

ここで、仕様とは、合成音声の品質レベル、音声情報データベースの容量、録音作業を含む音声情報データベースの作成時間等であり、これらの仕様に基づく要求を満たすように、文字列選択手段の動作が制御される。

【0015】

この発明はさらに追加録音原稿作成装置を提供している。この追加録音原稿作成装置は上記の録音原稿作成装置に付随するものでもよいし、独立したものでもよい。

【0016】

この追加録音原稿作成装置は、既存の音声情報データベースに含まれるすべての第1の音声単位を抽出する音声情報データベース分析手段、追加元原稿に含まれる文字列を構成す

50

るすべての第2の音声単位を抽出する追加元原稿分析手段、第2の音声単位について、第1の音声単位に含まれていない音声単位を検出する比較手段、および上記比較手段によって検出された音声単位を含む文字列を追加元原稿から選択して追加録音原稿を作成する第2の文字列選択手段を備えているものである。

【0017】

既に作成された音声情報データベースのデータを有効に利用し、(できるだけ)最小限の文字列を含む追加録音原稿で追加原稿の文字列を音声合成できる音声情報データが作成できるようになる。

【0018】

録音原稿作成装置(追加録音原稿作成装置を含む)は、一般的には、音声情報データベース作成装置の一部として位置づけられ、音声情報データベース作成装置および録音原稿作成装置はコンピュータシステムにより実現される。このコンピュータシステムを制御する録音原稿作成のためのプログラムは、所与の元原稿に含まれる文字列を構成するすべての音声単位を抽出し、かつそれらの音声単位について、その元原稿における出現回数を検出し、出現回数の少ない音声単位を含む文字列から順に、上記の抽出したすべての音声単位を網羅するまで、元原稿から文字列を選択して録音原稿に加えるようにコンピュータを制御するものである。

10

【0019】

さらに、この発明による録音原稿作成方法は、所与の元原稿に含まれる文字列を構成するすべての音声単位を抽出し、かつそれらの音声単位について、その元原稿における出現回数を検出し、出現回数の少ない音声単位を含む文字列から順に、上記の抽出したすべての音声単位を網羅するまで、元原稿から文字列を選択して録音原稿に加えるものである。

20

【0020】

この発明による追加録音原稿作成方法は、既存の音声情報データベースに含まれるすべての第1の音声単位を抽出し、追加元原稿に含まれる文字列を構成するすべての第2の音声単位を抽出し、第2の音声単位について、第1の音声単位に含まれていない音声単位を検出し、検出された音声単位を含む文字列を追加元原稿から選択して追加録音原稿を作成するものである。

【0021】

この発明による録音管理装置は、録音原稿に含まれる複数の文字列を、所与の表示トリガごとに順次表示する表示装置、表示装置に表示された文字列について話者から入力された音声信号を一時的に記憶する録音手段、上記音声信号を分析する音声信号分析手段、上記音声分析手段の分析結果に基づいて音声採否を判定し、採用と判定したときには上記録音手段に一時記憶された音声信号を音声波形データベースに格納するように制御するとともに表示トリガを上記表示装置に与える音声採否判定手段、および上記音声信号分析手段の分析結果または上記音声採否判定手段の判定結果に基づいて、話者に与えるべき指示情報を作成する話者管理手段を備えているものである。

30

【0022】

表示装置には、話者が声を出して読むべき(発音すべき)文字列が表示されるから、話者はこの表示にしたがって、発話していけばよい。話者に対して、音声分析結果に基づく指示情報が与えられるから、話者はこの指示にしたがって行動すればよい。指示情報には、音声分析結果(たとえば、声の高さ、音量、話速など)や、これらについての注意、休息指示等が含まれる。さらに、話者の発話した音声についての採否が判定され、採用と判定されたときのみ話者の音声データベースに格納されるので、良質の音声波形データを確保することができる。このようにして、音声の分析結果とその履歴情報に基づいて、録音原稿を読上げた話者の音声をデータベースに登録するか否かを判定したり、発話に関する指示情報を話者に対してフィードバックするので、録音作業を管理するための録音ディレクタが付き添わなくても、話者ひとりだけで、録音作業を進めることができ、かつ質の高い音声データを収録することができる。

40

【0023】

50

好ましい実施態様では、上記表示装置に表示される文字列を表わす合成音声信号を作成する音声合成手段、および音声合成手段によって作成された合成音声信号を出力する音声出力手段がさらに設けられる。

【0024】

適切な読上げ方を手本として、標準音声を出力することにより、話者が録音原稿の文字列を不適切に読上げることが防ぎ、録音音声の品質を向上できる。

【0025】

この発明による録音管理方法は、録音原稿に含まれる複数の文字列を、所与の表示トリガごとに順次表示し、表示された文字列について話者から入力された音声信号を一時的に記憶し、上記音声信号を分析し、上記分析結果に基づいて音声採否を判定し、採用と判定したときには一時記憶された音声信号を音声波形データベースに格納するとともに表示トリガを発生し、上記音声信号の分析結果または上記音声採否判定結果に基づいて、話者に与えるべき指示情報を作成して出力するものである。

10

【0026】

この発明による録音管理のためのプログラムは、録音原稿に含まれる複数の文字列を、所与の表示トリガごとに順次表示し、表示装置に表示された文字列について話者から入力された音声信号を分析し、分析結果に基づいて音声採否を判定し、採用と判定したときには、一時記憶された音声信号を音声波形データベースに格納するとともに表示トリガを発生し、上記音声信号の分析結果または上記音声採否判定結果に基づいて、話者に与えるべき指示情報を作成するようにコンピュータを制御するものである。

20

【0027】

この発明によるラベリング装置は、録音原稿中の文字列と、この文字列を発音することにより得られる音声波形データとの対応づけにより、音声波形データを音声単位ごとに区切り、音声単位を表わすラベルとその区切りを表わす時間情報とを含む第1のラベル情報を作成する第1のラベリング手段、および上記第1のラベリング手段によって作成された第1のラベル情報における時間情報を修正または無効化するラベリングエラー除去手段を備えているものである。

【0028】

上記ラベリングエラー除去手段は、一実施態様では、音声単位ごとに設けられた修正規則に基づいて時間情報を修正するものである。

30

【0029】

上記ラベリングエラー除去手段は、他の実施態様では、上記第1のラベリング手段とは異なる第2のラベリング手段によって上記録音原稿中の文字列について作成された第2のラベル情報に含まれる時間情報と、上記第1のラベル情報の対応する時間情報との差を算出し、この差が測定値を超えている場合に、その時間情報について無効化情報を付与するものである。

【0030】

上記ラベリングエラー除去手段は、さらに他の実施態様では、既に作成されているラベル情報について統計的手法により、音声単位ごとに継続時間の信頼区間を作成し、上記第1のラベル情報に含まれる時間情報から生成される音声単位ごとの継続時間を対応する信頼区間と比較し、継続時間が信頼区間外の場合に、その継続時間を生成した時間情報に無効化情報を付与するものである。

40

【0031】

この発明によると、作成されたラベル情報の時間情報について、その時間情報が適切なものかどうかのチェックが行われ、必要に応じて修正または無効化されるので、最終的に得られるラベル情報は信頼性の高いものとなる。

【0032】

この発明によるラベリング方法は、録音原稿中の文字列と、この文字列を発音することにより得られる音声波形データとの対応づけにより、音声波形データを音声単位ごとに区切り、音声単位を表わすラベルとその区切りを表わす時間情報とを含むラベル情報を作成し

50

、ラベリング手段によって作成されたラベル情報における時間情報を修正または無効化するものである。

【0033】

この発明によるラベリングのためのプログラムは、録音原稿中の文字列と、この文字列を発音することにより得られる音声波形データとの対応づけにより、音声波形データを音声単位ごとに区切り、音声単位を表わすラベルとその区切りを表わす時間情報とを含むラベル情報を作成し、ラベリング手段によって作成されたラベル情報における時間情報を修正または無効化するようにコンピュータを制御するものである。

【0034】

この発明による音声情報データベース作成システムは、音声合成により作成すべき文字列を含む元原稿から、音声情報データベースに音声情報を格納すべき文字列を、元原稿に含まれる文字列の音声単位を分析してできるだけ少ない文字列ですべての音声単位を含むように選択して録音原稿を作成する手段、上記録音原稿作成手段により作成された録音原稿に含まれる複数の文字列を、所与の表示トリガごとに順次表示する表示装置、表示装置に表示された文字列について話者から入力された音声信号を一時的に記憶する録音手段、上記音声信号を分析し、この分析結果に基づいて音声採否を判定し、採用と判定したときには上記録音手段に一時記憶された音声信号を音声波形データベースに格納するように制御するとともに表示トリガを上記表示装置に与え、さらに上記分析結果または上記音声採否判定結果に基づいて、話者に与えるべき指示情報を作成する録音管理装置、上記録音原稿作成手段により作成された録音原稿中の文字列と、上記音声波形データベースに格納された音声波形データとの対応づけにより、音声波形データを音声単位ごとに区切り、音声単位を表わすラベルとその区切りを表わす時間情報とを含む第1のラベル情報を作成するとともに、作成されたラベル情報における時間情報を修正または無効化するラベリング装置、上記音声波形データベースに格納された音声波形から特徴量を作成する特徴量作成手段、ならびに上記音声波形データベースに格納された音声波形データと、上記ラベリング装置により作成されたラベル情報および上記特徴量作成手段により作成された特徴量を含むインデックス情報とを対応づけて記憶する音声情報データベース作成手段を備えているものである。

【0035】

この発明による音声情報データベース作成方法は、音声合成により作成すべき文字列を含む元原稿から、音声情報データベースに音声情報を格納すべき文字列を、元原稿に含まれる文字列の音声単位を分析してできるだけ少ない文字列ですべての音声単位を含むように選択して録音原稿を作成し、作成された録音原稿に含まれる複数の文字列を、表示装置に、所与の表示トリガごとに順次表示し、表示装置に表示された文字列について話者から入力された音声信号を一時的に記憶し、上記音声信号を分析し、この分析結果に基づいて音声採否を判定し、採用と判定したときには一時記憶された音声信号を音声波形データベースに格納するように制御するとともに表示トリガを上記表示装置に与え、さらに上記分析結果または上記音声採否判定結果に基づいて、話者に与えるべき指示情報を作成し、作成された録音原稿中の文字列と、上記音声波形データベースに格納された音声波形データとの対応づけにより、音声波形データを音声単位ごとに区切り、音声単位を表わすラベルとその区切りを表わす時間情報とを含むラベル情報を作成するとともに、作成されたラベル情報における時間情報を修正または無効化し、上記音声波形データベースに格納された音声波形から特徴量を作成し、そして上記音声波形データベースに格納された音声波形データと、作成されたラベル情報および作成された特徴量を含むインデックス情報とを対応づけて音声情報データベースに格納するものである。

【0036】

この発明によると、専門的な知識を持たない一般的の利用者であっても、比較的容易に、比較的短時間で、比較的高品質の音声情報データベースを作成できる。したがって、波形接続型音声合成において、一般の利用者でも容易に所望の声で、自然な合成音声を作成でき、波形接続型の音声合成が広く一般に普及するようになることが期待される。

10

20

30

40

50

## 【0037】

この発明は、特に、波形接続型の音声合成で用いられる音声情報データベースの作成を対象としているが、その他の合成方式（波形重畳型など）のためのデータベースの作成にも適用することができる。さらに、音声合成用途以外でもこの発明により作成した音声データベースは、音声認識のための統計的音響モデル（HMM）の学習データや音声分析のための試料データとしても利用することができる。

## 【0038】

## 【実施例】

## （1）波形接続型音声合成

波形接続型音声合成は、多数（複数）の単語、句、節、文についての音声波形データをあらかじめ用意しておき、これらの音声波形データから必要な部分を切出し（切出された音声波形を波形素片という）、複数の波形素片を組合せて接続することによって、新たな単語、句、節または文を表わす合成音声の音声波形を作成するものである。あらかじめ用意する音声波形データを元波形データという。元波形データには後述するようにインデックス情報が付随し、元波形データとインデックス情報のセット（これを波形情報という）は音声情報データベースに格納される。音声合成のために元波形データから必要な部分を切出す単位が音声単位である。

10

## 【0039】

この明細書において、音声単位は、単語、音節、音素および分割音素を含む。単語とは意味の一つのまとまりを表し、文法上の働きをもつものとしての言語の最小単位である。例えば、「ねこが寝る」という文において、「ねこ（neko）」、「が（ga）」、「寝る（neru）」はそれぞれ単語である。音節とは言語学上の発音の単位である。例えば、「ね（ne）」、「こ（ko）」などである。日本語ではかな文字の1つ1つが音節に相当し、100～300種類程度ある。音節は1つまたは複数の音素で構成される。音素とは、音声の基本的な最小単位である。例えば、「n」、「e」、「k」、「o」などである。音素は、母音（Vowel、記号Vで表す）と子音（Consonant、記号Cで表す）に分類される。日本語では、母音は5種類（a、i、u、e、o）、子音は約20種類（n、k、s、t、m、rなど）がある。分割音素とは、音素をさらに分割したものであり、いくつに分割したものは問わない。音素は、波形接続型音声合成において最も一般的に用いられる音声単位である。音節も、一般的に用いられる音声単位の1つである。

20

30

## 【0040】

以上に基づいて、「音声単位」を次のように定義する。すなわち、音声単位とは、母音または子音である音素を分割した分割音素を1つまたは連続させたものである。換言すれば、すべての音声単位は、1つまたは連続する複数の分割音素により構成される。

## 【0041】

波形接続型の音声合成では、音声単位として、音節や音素のほかに、VCV素片やCVC素片などの音韻環境を考慮した音声単位も一般的に用いられる。音韻環境を考慮した音声単位とは、ある音声単位について、その前後（両方またはいずれか一方）の音声単位の違いも含めて種類を区別したものである。上では連続する3つの音素からなる音声単位（VCV素片、CVC素片）を2種類挙げているが、これ以外にも連続する1つ以上の音節からなるものや、連続する1つ以上の分割音素からなるものなど、音韻環境を考慮した音声単位には様々なものが存在する。VCV素片とは、母音、子音、母音の3つの連続する音素を1つの音声単位とみなしたものである。例えば、「e-k-o」や「o-g-a」などで種類は700～800程度ある。CVC素片とは、子音、母音、子音の3つの連続する音素を1つの音声単位とみなしたものである。例えば、「n-e-k」や「k-o-g」などで種類は5000～6000程度ある。

40

## 【0042】

図1は音声波形において、音素、音節および単語の区切りをつけて、音声波形と対応付けて示すものである。図2は、音韻環境を考慮した音声単位を音声波形と対応付けて示すものである。

50

## 【 0 0 4 3 】

音声波形とは、空気の振動（音）により発生する空気の粗密を時間変化として表したものである。図 1 および図 2 のような音声波形の図において、横軸は時間を、縦軸は空気の密度の高さをそれぞれ表している。音声波形をコンピュータ上で扱うときには通常サンプリング処理により標準化された時系列データを音声波形ファイルとして取扱い、音声波形ファイルに録音（ファイル保存）、書込み、読出し等の処理を行う。音声波形データの開始時点からの経過時間を用いて各音声単位の始点、終点および継続時間を表すことができる。

## 【 0 0 4 4 】

図 1 において、音声波形の開始点から音の始まりに相当する区間には、ポーズ（無音）を示すラベル「pau」が与えられ（「ラベル」については後に説明する）、「t」、「a」、「n」、「a」、「k」、「a」の音声単位（音素）で音声波形が区切られている。音声波形の下段には、音素の区切り、音節の区切りおよび単語の区切りが音声波形に対応付けて示されている。

10

## 【 0 0 4 5 】

図 2 は、子音「k」を中心として、前後の母音「a」を音韻環境として考慮した V C V 形式の音声単位の音声波形を示すものである。この音声波形において、上段には音素単位で音韻環境を考慮した音声単位が示され、下段には分割音素単位で音韻環境を考慮した音声単位が示されている（ここでは、分割音素は 1 音素を 2 つに分割したものである）。図 2 の下段において、音素「a」の前半の分割音素を「a |」で後半の分割音素を「| a」で表している。

20

## 【 0 0 4 6 】

図 3 は音声波形とラベル情報との関係を示す。

## 【 0 0 4 7 】

ラベル情報は、音声波形を音声単位で区切ったときの音声単位（音声波形を構成する音声単位）ごとに設けられ、その音声単位における符号（これをラベルという）（たとえば、音声単位が音素の場合には、n, e, k, o 等のアルファベット、音声単位が音節の場合には n e, k o などのかな文字）と、その音声単位の音声波形における時間位置情報（単に、時間情報という）とから構成される。時間情報は、音声単位の終りの位置がどこであるのか（音声単位の終点）または音声単位の始まりの位置がどこであるのか（音声単位の始点）を示す情報である。

30

## 【 0 0 4 8 】

コンピュータ上では、ラベル情報を、各音声単位を表すラベル（アルファベット記号で表記）とその終点を表す時間情報の組を時系列順に記述したテキスト・ファイルとして扱う。この場合、各音声単位の始点は、直前の音声単位の終点に等しく、各音声単位の継続時間は、直前の音声単位の終点を示す時間情報とその音声単位の終点を示す時間情報との差によって求めることができる。音声波形ファイルの開始時点から音の始まりまでに相当する区間には、ポーズ（無音）を示すラベル「pau」が与えられる。音声波形ファイルの録音時に、正確に音の開始点と終了点で録音を開始、停止することが難しいため、通常は、音声波形の先頭や末尾にはポーズが含まれる。図 3 において、音声単位の終点（0.160, 0.250 など）を時間情報として保持するためには、音の始点を示すために先頭のポーズの終点情報（0.120）が必要である（逆に音の終点は最後の音声単位の終点に等しいため、末尾のポーズの終点を示す時間情報は必要ない）。

40

## 【 0 0 4 9 】

上述したように、波形情報データベースには複数の音声波形についての波形情報が格納される。波形情報は音声波形データとインデックス情報とから構成される。インデックス情報とは、音声波形（元波形）ごとに、その音声波形を構成する各音声単位について、ラベル情報と音声波形の特徴量（音声単位ごと）とを記述したものである。

## 【 0 0 5 0 】

特徴量には、音声波形（音声単位ごと）の音韻的特徴と韻律的特徴がある。音韻的特徴に

50



は、ケプストラムおよびベクトル量子化データが含まれる。ケプストラムは、音声波形の短時間振幅スペクトルの対数を逆フーリエ変換したものである。ベクトル量子化データは、音声波形の複数のパラメータ値のベクトルを代表ベクトルの符号で表したものである。また韻律的特徴には基本周波数、パワーおよび上述した継続時間が含まれている。基本周波数とは、音源である声帯が振動する周波数であり、音声の「高さ」(ピッチ)を表す指数である。基本周波数が高いほど声の高さは高くなる。パワーとは、音声波形の振幅である。音の「大きさ」に対応する。継続時間は、換言すると音声単位に相当する音声波形の時間長(「長さ」)である。音声単位の長さに対応する音声の継続時間(一つの音声波形で考えると、継続時間の平均値)が小さいことは話速が速いことを示す。

【0051】

10

図4は波形情報(音声波形データとインデックス情報のセット)を用いて波形接続型音声合成を行う様子を示している。「さかた」と発音(発話)する合成音声の音声波形を作成するために、「さとう」と発話された音声波形(これを音声1とする)と「たなか」と発話された音声波形(これを音声2とする)の2つの元波形を用いる。これらの元波形を含む音声情報データベースに格納されたインデックス情報が図4の左側に示されている。音声(音声1, 2を含む)のそれぞれについて、インデックス情報は、各音声波形を構成する音声単位(ここでは音素)のラベルおよび始点(以下、ラベル情報)と、長さ(時間長)、高さ(周波数)および大きさ(振幅)(以下、波形の特徴量)を含む。

【0052】

作成すべき合成音声を表わす文字列「さかた」が与えられると、インデックス情報を参照して、「sakata」の音声波形を合成するのに必要な音声単位を選択する。音声1から「s」および「a」、音声2から「t」および「a」と「k」および「a」がそれぞれ選択される。

20

【0053】

選択された各音声単位に対応する波形素片を、インデックス情報に記述された始点と長さに基づいて、元波形からそれぞれ切出す。音声1の元波形から「s」と「a」をそれぞれ表わす波形素片が、音声2の元波形から「t」と「a」をそれぞれ表わす波形素片と「k」と「a」をそれぞれ表わす波形素片がそれぞれ切出される。これらの波形素片が「s」、「a」、「k」、「a」、「t」、「a」の順序に接続(合成)される。

【0054】

30

このように、元波形から切り出した波形素片に対して、信号処理を行うことなく、波形素片を所与の順序で接続するので、音質を劣化させることなく合成音声の音声波形を作成することができる。

【0055】

図5は波形接続型音声合成処理の流れを示すものである。

【0056】

音声合成により作成すべき発音(発話)を表わす文字列が与えられる。この入力文字列は、音声単位のラベル列に変換される。例えば、日本語の場合に漢字かな交じりの文の入力があつたとすると、この文の単語への分割、幾つかの単語をグループ化したうえでアクセント位置の決定、単語グループ間に挿入するポーズ(間)の長さの決定などの処理を行う

40

【0057】

韻律予測処理92では、音声単位ラベル列に基づいて、各音声単位の韻律的特徴を予測する。具体的には、音声情報作成における特徴量抽出の処理で、音声単位ごとに音の高さ、強さ、長さのパターンを抽出した結果を利用する。韻律的特徴を直接指定して入力してもよい。

【0058】

音声単位選択処理93では、音声情報データベース97から音声単位ラベル列のラベルと一致する音声単位を選択する。一致する音声単位が複数ある場合には、音声情報データベースのインデックス情報を参照して、韻律的特徴が最も一致する音声単位を選ぶようにする。

50

## 【 0 0 5 9 】

波形接続処理94では、選択された音声単位のインデックス情報を参照して、元波形データからその音声単位に相当する波形素片を切出し（信号処理せずそのまま）、音声単位ラベル列の順に接続する。

## 【 0 0 6 0 】

音声出力処理95では、接続して出来上がった合成音声の音声波形を音声デバイス（たとえばスピーカ）96へ送り、音を出力する。

## 【 0 0 6 1 】

波形接続型音声合成は、音声波形データに対して信号処理を行わないという特徴をもつため、以下のような長所がある。

- ・信号処理による音質の劣化がない。一般に音声波形に対して信号処理を行うと、声が不自然になるなど音質の劣化が発生する。
- ・元の音声波形データの声の特徴をそのまま残した合成音声を得られる。特定の人物たとえばアナウンサやタレントなどと同じ声の特徴を持つ合成音声を作成できる。
- ・音声情報データベースを交換することにより合成音声の声を自由に変えられる。

## 【 0 0 6 2 】

また、予め用意した音声波形をもとにして合成音声を作成するために以下の点を考慮しなければならない。

- ・合成したい音の全てを含むような音声波形データ（元波形データ）を用意し、かつ元波形データの量が大きくなりすぎないようにする。すなわち元波形として用意されていない音は合成できない。また、元波形のデータの量が大きくなりすぎると音声情報データベースに入らない。
- ・十分に良好な音質の元波形データを用意し、かつ元波形データに音質のばらつきがないようにする。
- ・元波形データから必要な部分を探し出して切り出すために、元波形の内容を示す情報（インデックス情報）を作成する必要がある。

## 【 0 0 6 3 】

## ( 2 ) 第 1 実施例

図6は音声情報データベース作成システムのハードウェア構成を示すブロック図である。このシステムは最も典型的には、いわゆるパーソナル・コンピュータまたはワークステーションとその周辺機器により実現することができるが、もちろん、音声情報データベース作成システム専用のハードウェア・アーキテクチャを持つものでよい。

## 【 0 0 6 4 】

音声情報データベース作成システムは、演算装置（CPU）20、ワークメモリ（RAM）21、通信I/F部22、入力I/F部23、出力I/F部24、データベース25、画面データメモリ26、処理プログラムメモリ27、入力装置28、出力装置29および合成音声出力装置30を含んでいる。

## 【 0 0 6 5 】

演算装置20は、音声情報データベース作成処理、その他のシステム管理処理のためのプログラムを実行する。

## 【 0 0 6 6 】

ワークメモリ21は、音声情報データベース作成処理における入出力データや中間処理データを格納するためのメモリである。

## 【 0 0 6 7 】

通信I/F部22は、入出力装置等のハードウェアを接続する場合、または外部機器と直接またはネットワークを介して通信するためのものであり、ノイズ除去や同期処理などを実行する。ネットワークは用途に応じて適切なものを使用すればよい。

## 【 0 0 6 8 】

データベース25は、音声情報データベース作成システムにおいて作成された各種データベース（詳細は後述する）を格納するためのものである。

10

20

30

40

50

## 【 0 0 6 9 】

画面データメモリ26は、出力装置に含まれる画面表示装置に出力される画面データを保持するメモリである。

## 【 0 0 7 0 】

処理プログラムメモリ27は、音声情報データベース作成処理のための各種実行プログラム（OSを含む）（このプログラムの詳細については後述する）を格納するメモリである。上述した各種メモリは、半導体メモリ、磁気ディスク、光ディスク、光磁気ディスク、その他の記憶媒体により実現される。

## 【 0 0 7 1 】

入力装置28は、操作者が音声情報データベース作成システムに情報を入力する為のものであり、例えば、キーボード、マウス、マイクロフォン、FDドライブ、表示画面等を含むものであり、入力I/F23を介して演算装置20と接続される。

10

## 【 0 0 7 2 】

出力装置29は、音声情報データベース作成システムの操作者に情報を出力するものであり、例えば、ディスプレイ（表示装置）、スピーカ等の操作者に情報を伝達するものであり、出力I/F24を介して演算装置20と接続される。

## 【 0 0 7 3 】

この音声情報データベース作成システムが、作成した音声情報データベースを用いて所望の音声を合成する機能（図6に示す）を持つ場合には、合成した音声を表わす波形データは合成音声出力装置30により記録媒体31に記録される。記録媒体は、CD-ROM、フロッピーディスク、DVD等を含む。

20

## 【 0 0 7 4 】

図7は上記の音声情報データベース作成システムにおいて、主に演算装置20が達成する諸機能を幾つかにまとめて表す機能ブロック図である。

## 【 0 0 7 5 】

この音声情報データベース作成システムには4つのデータベース、すなわち原稿データベース11、音声波形データベース12、ラベル情報データベース13、および最終的に作成されるべき音声情報データベース15が含まれる。これらのデータベースは基本的にこのシステムが運用される過程で作成されるもので、具体的には図6に示すデータベース25に対応する。

30

## 【 0 0 7 6 】

仕様入力部（手段）4は、この音声情報データベース作成システムを運用する操作者OPが音声情報データベースを作成するにあたって定める仕様（事項）（音声情報データベース容量、音声情報データベース品質、作成時間および元原稿ファイル名）を入力（コンピュータに取込む）するもので、具体的には、図6に示す入力装置28により実現され、詳細は図8に示されている。

## 【 0 0 7 7 】

原稿作成部（手段）5は、仕様入力部4から入力された仕様情報に応じて原稿データベース11内の元原稿、または仕様入力部4から与えられる元原稿に基づいて録音原稿を作成するものである。録音原稿とは、話者SPが声に出して読む原稿（すなわち、録音されるべき原稿）をいう。話者SP（話し手、発話者）は録音原稿を声に出して読む人である。システムの操作者OPと話者SPとは同一人でも、異なる人でもよい。原稿作成部（手段）5は、図6に示す処理プログラムメモリ27に格納された原稿作成プログラム（図11参照）を実行する演算装置20により実現され、詳細については図8を参照して後述する。

40

## 【 0 0 7 8 】

録音管理部（手段）6は、話者SPの発話音声（または録音音声）の分析結果とその履歴情報に基づいて、その音声を音声情報データベースに収録すべきかどうかの判定、話者SPに対する発話の指示、長時間にわたる録音作業の過程で不可欠な休息時間の設定等を行うものである。これにより、録音ディレクタ（操作OP）の付き添いがなくても、話者SPだけで録音作業を進めることができ、かつ質の高い音声波形データを収録することが可

50

能になる。録音管理部（手段）6は、処理プログラムメモリ27内の録音管理プログラム（図15, 16参照）とこれに従う動作を行う演算装置20とにより実現され、その詳細は図9に示されている。

【0079】

表示装置9は、原稿作成部（手段）5によって作成された原稿の表示、録音管理部（手段）6から出力される休息指示、発話注意等の表示を行うもので、図6の出力装置29に含まれる。

【0080】

音声入力装置（手段）10は、話者が発生する音声（発話音声）を電気信号（音声波形）に変換するもので、マイクロフォンにより実現される。図6の入力装置28に含まれる。

10

【0081】

録音部（手段）7は、音声入力装置10から入力する音声波形に基づいて発話開始および終了を検出するとともに、検出した発話開始と終了との間の音声波形を記録媒体（磁気テープ、磁気ディスク、半導体メモリ等）に一時的に記録する。音声波形は好ましくはデジタルデータに変換されるがアナログのまま一時的に保持してもよい。録音部7の詳細は図9に示され、図6の入力I/F23に対応する。

【0082】

ラベリング部（手段）8は、録音原稿作成部（手段）5で作成された録音原稿を記録した音声波形データのラベル情報を作成する。さらに、作成したラベル情報からラベリングエラーを検出し、ラベリングエラー箇所の修正または除去を行う。これにより、熟練者の技能を必要とせず熟練者と同じ水準でラベリング情報を作成することができる。ラベリング部（手段）8は、図6に示す処理プログラムメモリ27に格納されたラベリングエラー除去プログラム（図18参照）を実行する演算装置20により実現され、詳細については図10を参照して後述する。

20

【0083】

特徴量抽出部（手段）4は、ラベル情報を参照しながら、音声波形ごと、または音声単位ごとに音律または音韻特徴を算出し、音声情報データベース15のインデックス情報を作成する。特徴量抽出部14は、処理プログラムメモリ27内の特徴量抽出プログラムとこれに従う動作を行う演算装置20とにより実現される。

【0084】

出力装置16は、音声情報データベース15に記録された音声情報をCD-ROM、フロッピーディスク、DVD等の記録媒体17に記録するものである。

30

【0085】

操作者OPは、作成すべき音声情報データベースに関する仕様を仕様入力部4を用いて入力する。仕様入力部4は、図8に示すように、FDドライブ（記録媒体読取装置）41と入力装置42を含む。入力装置42は、図12に示すような仕様入力画面を表示する表示装置、表示画面上のボックス等に文字、数字等を入力するためのキーボード、各種操作のマウス等を含む。

【0086】

仕様の項目には、作成すべき音声情報データベースの上限容量、同データベースの品質、同データベースを作成するのに要する（許容できる）上限作成時間、および元原稿ファイル名がある。上限容量は、一般的に動作環境やアプリケーションのデータ領域の制限上、音声情報データベースのために使用できるメモリ容量が制限される場合に用いる。品質は、高いほど音声情報データベース15の容量は大きくなるが合成音声の品質も高くなる（詳細は後述する）。作成時間は主に話者SPが音声を入力作業を行う時間である。

40

【0087】

音声情報データベースの作成時間が長ければ、データベースの容量は増大する。したがって、上限作成時間はデータベース容量を制限する。データベース作成時間は作成されるデータベースの容量に比例すると考えて良いので、入力された上限作成時間をデータベース容量に次式を用いて変換することができる。

50

## 【0088】

データベース容量 = データベース作成時間 × 変換係数

## 【0089】

変換係数は、データベース作成時間とデータベース容量の比を示す値で、予め用意しておく、または実績値に基づいて調整することが可能である。すなわち、実際の音声情報データベースの作成終了時点において、完成した音声情報データベースの容量と作成に要した時間に基づいて次式を用いて変換係数を調整する。

## 【0090】

**調整後の変換係数 = (1 - 調整重み) × 調整前変換係数  
+ 調整重み × (データベース容量 / データベース作成時間)**

10

## 【0091】

音声情報データベースの品質は整数値で表されるレベルで表記される。品質レベルが高くなるほど音声単位の種類は増え、その音声情報データベースを用いて生成される合成音声の質も高くなる。この実施例では品質レベルは3レベルあり、例えば、元原稿中のすべての音素が含まれているという品質が「レベル1」、すべての音節が含まれているという品質が「レベル2」、アクセントの有無を区別した音節が含まれているという品質が「レベル3」である。例えば「すずき (suzuki)」という音声は、レベル1ではs, u, z, k, iの5種類、レベル2では, su, zu, kiの3種類の単位にそれぞれ分類される。品質レベルが高くなれば、データベース容量が増大し、作成時間が長くなる。元原稿ファイル名

20

## 【0092】

操作者OPが音声情報データベースの仕様を入力する場合に、音声情報データ作成システムの表示装置に、図12に示す仕様入力画面が表示される。

## 【0093】

この仕様入力画面の左端には、開始、仕様入力、原稿作成、録音、ラベリング、特徴量抽出、終了の順に音声情報データベース作成の工程が表示され、現在行っている工程の表示に、周囲や他の工程とは異なる色が付される。画面上段に表示された仕様入力領域には、音声情報データベースの容量 (DB容量)、同データベースの品質 (DB品質) レベル、作成時間の各希望値を入力するボックスと、原稿ファイル名を入力するボックスとがある。さらに、入力を確定する「設定」ボタンが設けられている。画面下段に表示された完成時の音声情報データベースの属性表示領域には、DB容量、DB品質レベルおよび作成時間について、予め設定されたデフォルト値と、操作者OPが入力した設定仕様値が表示される。

30

## 【0094】

この仕様入力画面において入力されるDB容量、DB品質および作成時間は仕様入力部42から原稿作成部5の文字列選択処理53に与えられる。DB容量と作成時間については少なくともいずれか一方が入力されていればよい。

## 【0095】

仕様入力画面において元原稿ファイル名が入力されていれば、その入力ファイル名は入力装置42からFDドライブ41に与えられる。FDドライブ41は装着されたFDに格納されているファイルのうち、入力された元原稿ファイル名の元原稿ファイルを読み出し、原稿作成部5の元原稿設定処理51に与える。

40

## 【0096】

図8において原稿作成部5は、元原稿設定処理 (手段) 51、元原稿分析処理 (手段) 52および文字列選択処理 (手段) 53を含んでいる。これらの各処理の動作を図11を参照して説明する。

## 【0097】

原稿作成部5は仕様入力部4から仕様データが与えられると、原稿作成処理を開始する (ステップS1)。

50

## 【 0 0 9 8 】

元原稿設定処理51はF Dドライブ41から元原稿ファイルが与えられているかどうかを判断する(ステップS 2)。元原稿ファイルが与えられていればその元原稿ファイルをワークエリアに取込む(ステップS 3)。元原稿ファイルが与えられていない場合には、元原稿設定処理51は原稿データベース11から既存の元原稿ファイルを読み出し、読み出した元原稿ファイルをワークエリアに設定する(ステップS 4)。

## 【 0 0 9 9 】

原稿データベース11に複数の元原稿ファイル(既に作成されて格納されているもの)が存在する場合には、仕様情報に含まれるD B容量およびD B品質に基づいて適切なものを選択するようにしてもよい。また、F D等の記録媒体から読み出した元原稿ファイルと原稿データベース11から読み出した元原稿ファイルとを組合わせたものを元原稿として設定してもよい。元原稿(元原稿ファイル)とは録音原稿の元(源)になる単語、句、節、文等を格納したもので、この元原稿から所要の単語、句、節、文を取出して後述するように録音原稿が作成される。

10

## 【 0 1 0 0 】

元原稿分析処理52はワークエリアに設定された元原稿に含まれる文字列を分析して、文字列を構成する各音声単位が元原稿に出現する回数を計測する(ステップS 5)。

## 【 0 1 0 1 】

図13(A)は元原稿の一例を示すものである。この元原稿は日本人の多くの苗字を列挙したものである(図では一部のみが示されている)。この元原稿は各苗字を表わす文字列のリストである。

20

## 【 0 1 0 2 】

このような元原稿が分析される。分析とは、品質レベルに応じて元原稿に記述された単語、句、節、文などを音声単位に分解することである。この実施例では、品質レベル1の音声単位は音素、品質レベル2の音声単位は音節、品質レベル3の音声単位はアクセントを含む音節である。設定されている品質レベル以下の品質レベルのすべてについて、それぞれに応じた音声単位への分析が行なわれる。品質レベル3が設定されているとすると、品質レベル1における音素への分解、品質レベル2における音節への分解および品質レベル3におけるアクセントを含む音節への分解のすべてが行なわれる。

## 【 0 1 0 3 】

このように分解された音声単位のすべてについて、品質レベル別に各音声単位が元原稿中に出現する回数を計測し、元原稿分析結果として音声単位リストを作成する。図7(B)は元原稿分析結果を示すものである。元原稿分析結果は品質レベル別の音声単位リストとして記述され、このリストでは出現回数の小さい順に配列され、出現回数と同じものについてはアルファベット順に並べられる。母音だけからなる音節は音素であり、品質レベル1のものとしてリストアップされているから、品質レベル2および品質レベル3のリストには含まれていない。

30

## 【 0 1 0 4 】

原稿作成部5における文字列選択処理53は、元原稿に含まれる単語、句、節、文章(これらを文字列という)に基づいて、先に作成した元原稿分析結果を参照して、できるだけ少ない文字列で、できるだけ多くの音声単位を含むような録音原稿を作成するものである。このために、録音原稿に加えるべき文字列を元原稿から次のように選択する。すなわち、まず最も低い品質レベルについての元原稿分析結果リストを参照して、最も出現回数の少ない音声単位を含む文字列(苗字)を元原稿から選択し、録音原稿に移す(追加する)(ステップS 8)。録音原稿に追加した文字列に含まれるすべての音声単位を元原稿分析結果リストから削除する(ステップS 9)。さらに選択した文字列を元原稿から削除する(ステップS 10)。元原稿分析結果リストにおいて出現回数が少ない音声単位の順に元原稿分析結果リストに残っている音声単位が無くなるまで、上記の処理を繰返す(ステップS 7)。

40

## 【 0 1 0 5 】

50

最も低い品質レベルについて、終了すれば、次の品質レベルの元原稿分析結果リストを参照して、録音原稿に追加すべき（移すべき）文字列（苗字）を元原稿において選択する。この処理は設定された品質レベルに達するまで繰返される。

【0106】

図14(A)は品質レベル1について作成された録音原稿の例を示している。この録音原稿では、4つの苗字が列挙されている。この4つの苗字は、図13(B)に示す品質レベル1についての元原稿分析結果リストのすべての音声単位を含んでいる。

【0107】

図14(B)は品質レベル2についての処理が終了した時点で得られる録音原稿の例を示している。図14(A)の録音原稿と比較すると2つの苗字（しみず、みやもと）が追加されている。これは、図13(B)に示す品質レベル2についての元原稿分析結果リストに挙げられた音声単位（音節）のすべてを含むように苗字を追加的に選択したことによる。

10

【0108】

品質レベル3が設定されている場合には、さらに品質レベル3の要求を満たす文字列の選択と追加が行なわれ、図14(C)に示すような録音原稿が得られる。これは図13(B)に示す品質レベル3についての元原稿分析結果リストに挙げられているアクセントを含む音節のすべてを含むように、元原稿から苗字を抽出したことによる。

【0109】

仕様入力部4において、上述したように音声情報DB容量、DB品質および作成時間が入力される。このうち、要求されたDB品質（品質レベル1～3）を満たすように上述の処理が行なわれる。すなわち、要求されたDB品質が品質レベル2であれば、図14(B)の録音原稿が得られた時点で処理が終了し、品質レベル3が要求されている場合には図14(C)の録音原稿が得られるまで処理が続けられる。

20

【0110】

他方、要求されたDB容量および作成時間もステップS8～S10の処理の繰返しを制御するために用いられる。作成時間は上述したようにDB容量に換算できる。仕様入力部4において入力されたDB容量、または入力された作成時間から換算されたDB容量のうちのいずれか小さい方がワークエリアに設定される（ステップS6）。元原稿から文字列（苗字）が選択され、録音原稿にその選択された文字列が移される（加えられる）たびに、加えられた文字列（苗字）についての音声情報容量（音声情報データベース15に格納される波形データ等を含むデータ容量）がワークエリアのDB容量から減算される。この減算結果を残りDB容量という。残りDB容量が零になると録音原稿作成処理は、たとえ途中であっても、終了する（ステップS7）。

30

【0111】

図7において、原稿作成部5において上述のように作成された録音原稿は録音管理部6に与えられる。録音管理部6では、後述するように録音原稿に含まれる文字列（苗字）を順次表示装置9に表示させるとともに、必要に応じて休息指示および発話注意を生成して表示させる。

【0112】

話者SPは、表示装置9に表示された文字列を表示の順序にしたがって声を出して読む（発話する）。

40

【0113】

話者SPにより発話された音声は音声入力装置10に入力され、電気信号に変換される。

【0114】

音声入力装置10から出力される音声を表わす電気信号は音声波形信号として録音部7および録音管理部6に入力される。録音部7に入力された音声波形信号は音声波形データとして録音（保存）される。録音管理部6は、後述するように入力された音声波形を分析する。分析の結果、良品質の音声波形であると判定した場合には、録音管理部6は、録音部7に音声波形データを音声波形データベース12に保存させる指令を与える。

【0115】

50

録音管理部 6 は、機能の観点から大きく分けると、話者管理処理（手段）6 a、音声分析処理（手段）6 b、音声採否判定処理（手段）6 c および録音管理処理（手段）6 d を備えている。話者管理処理（手段）6 a は発話注意生成処理（手段）61、休息指示生成処理（手段）62、音声分析結果保持処理（手段）63 を備えている。音声分析処理（手段）6 b は、基本周波数検出処理（手段）64、音量検出処理（手段）65、話速検出処理（手段）66 を備えている。音声採否判定処理（手段）6 c は音声分析結果比較処理（手段）67 および音声採否判定処理（手段）68 を備えている。

【0116】

録音部 7 は、発話開始、終了検出処理（手段）71 および録音処理（手段）72 を含んでいる。

10

【0117】

話者は表示装置 9 の表示にしたがって、録音原稿内の文字列（苗字）を一つずつ声を出して読む。一つの文字列についての音声信号が音声入力装置 10 から録音管理部 6 および録音部 7 に与えられる。

【0118】

音声分析処理 6 b は、音声入力装置 10 から入力された一つの文字列の音声信号について、その基本周波数（高さ）、音量（パワー）および話速をそれぞれ処理 64、65、66 において検出し、これらの検出結果を音声波形分析結果として音声採否判定処理 6 c の音声分析結果比較処理 67 と話者管理処理 6 a の音声分析結果保持処理 63 に与える。

【0119】

音声採否判定処理 6 c の音声分析結果比較処理 67 は、予め設定して音声波形データベース 12 に記憶しておいた音声採否判定基準を読み出し、与えられる音声波形分析結果と読み出した音声採否判定基準とを比較して、音声入力装置 10 から録音部 7 に入力された音声を音声波形データベース 12 に音声波形データとして登録するか否かの判定を行う。音声波形分析結果の各属性（基本周波数、音量、話速）の全てが音声採否判定基準の範囲内に収まっている場合には、録音部 7 に保存された音声波形データを音声波形データベース 12 に保存させ（採用と決定）、それ以外の場合には、録音部 7 に音声波形データを消去（不採用と決定）させる。この動作は各文字列を表わす音声信号について順次行なわれる。

20

【0120】

音声分析結果保持処理 63 は、音声分析処理 6 b から出力された音声波形分析結果の履歴情報を保存しておく。また、音声分析結果保持処理 63 は音声採否判定処理 68 による採否判定結果を受取る。採否判定結果が不採用であったときには、音声分析結果保持処理 63 は録音管理処理 6 d に繰返し指令を与え、不採用となった音声に対応する文字列を再度表示装置 9 に表示させる。

30

【0121】

発話注意生成処理 61 または休息指示生成処理 62 は、音声分析結果保持処理 63 に保持された音声波形分析結果の履歴情報または、採否判定結果についての情報に基づいて、次のようにして、必要に応じて、発話注意または休息指示を生成して、録音管理処理 6 d に与える。

【0122】

発話注意生成処理 61 は、波形分析結果（周波数、音量、話速）について平均値を常時算出している。そして、今回の波形分析結果とこの平均値を比較し、比較結果に応じて、発話注意を生成する。たとえば、今回の音量と音量の平均値とを比較し、今回の音量が音量の平均値を大きく下回っていれば（差が所定の閾値以上であれば）、「声が小さくなっています」という発話注意を生成する。

40

【0123】

休息指示生成処理 62 は音声採否判定処理 68 が不採用と判定した頻度に基づいて休息指示を発生する。たとえば、今回の不採用判定が前回の不採用判定に近ければ、話者の疲れが原因で不採用が頻発していると考えられるので、休息指示を発生する。

【0124】

50



録音管理処理 6 d は、原稿作成部 5 から与えられる録音原稿を保持し、順次表示装置 9 に発話すべき文字列を表示する。表示装置 9 に表示される画面の一例が図 17 に示されている。この画面では 31 番目の文字列（苗字）として「佐藤」が表示されている。

【 0 1 2 5 】

音声採否判定処理 68 の採否判定結果は音声分析結果保持処理 63 を介して録音管理処理 6 d に与えられるので、録音管理処理 6 d は採用判定であれば次の文字列（苗字）を表示装置 9 に表示させ、不採用であれば前回と同じ文字列（苗字）を表示させるように表示装置 9 を制御する。

【 0 1 2 6 】

録音管理処理 6 d はまた、発話注意生成処理 61 から与えられる発話注意や休息指示生成処理 62 から与えられる休息指示を表示装置 9 に表示させるように制御する。図 17 の表示画面では、アドバイスとして、「10 分間の休憩をとってください」という休息指示と、「声が小さくなっています」という発話注意とが表示されている。

【 0 1 2 7 】

表示装置 9 にはまた、発話注意処理 61 が算出した音声分析結果の平均値（ハッチングで示す）と今回の音声分析結果が音声の属性ごとに（音量、話速、高さ、発話内容）グラフで表示されている。発話内容は、音声認識による信頼度を示すスコアである。

【 0 1 2 8 】

休息指示生成処理 62 は休息指示を出力した後、指示した休息時間が経過したときに再開指示を録音管理処理 6 d に与える。録音管理処理 6 d はこれに応答して、発話すべき文字列の表示を続ける。

【 0 1 2 9 】

なお、図 17 において、「録音」ボタンは話者が発話の開始を明示的に入力する場合に用いるもので、発話開始検出機能が備えられている場合には不要である。「再生」ボタンは話者が録音音声を再生して確認するとき用いるものである。

【 0 1 3 0 】

録音部 7 には音声入力装置 10 からの音声信号が入力している。発話開始・終了検出処理 71 は入力する音声信号の開始時点と終了時点を検出するもので、これらの開始時点から終了時点までの間の音声信号が録音装置 72 に与えられて録音される。

【 0 1 3 1 】

図 15 および図 16 は、録音管理部 6 による録音管理処理を示すフローチャートである。

【 0 1 3 2 】

録音管理処理 6 d は、原稿作成部 5 によって作成された録音原稿を読み込む（ステップ S 21）。このとき、録音済みの文字列（苗字）の数（録音済件数）（変数またはカウンタ）を 0 にリセットし、録音原稿の文字列数（録音原稿に含まれる文字列（苗字）の総数）を、録音全件数（変数またはカウンタ）としてセットする（ステップ S 22）。

【 0 1 3 3 】

録音管理部 6 d は、録音済件数が録音全件数よりも小さいか否かの判定を行う（ステップ S 23）。録音済件数が録音全件数以上になった場合には、録音処理を終了する（ステップ S 23 で No）。

【 0 1 3 4 】

録音済件数が録音全件数よりも小さい場合には、録音管理処理 6 d は、録音原稿の文字列リストの中から（録音済件数 + 1）番目の文字列を読み上げ文字列として設定し（たとえばバッファに格納し）（ステップ S 24）、これを表示装置 7 に出力する（ステップ S 25）。

【 0 1 3 5 】

表示装置 7 には、図 17 に示すような録音表示画面が表示される。上述した仕様入力画面と同様に画面左側に音声単位データベース作成工程が表示されている。この段階では「録音」が明示されている。画面上段には録音原稿文字列表示領域があり、この領域には話者が読み上げるべき文字列（「佐藤（さとう）」）が表示される。画面中段には、上述したように音声波形分析結果領域がある。

10

20

30

40

50

## 【0136】

話者SPが読上げる文字列を発話する（声を出して読む）と、その音声が入力装置10に入力され、音声は入力装置10から音声波形として録音部7と音声管理部6の音声分析処理6bに入力する（ステップS26でYES）。録音部7に入力された音声波形は音声波形データとして録音される。

## 【0137】

音声分析処理6bは上述のように入力された音声波形を、高さ（基本周波数）、大きさ（パワー）、速さ（継続時間）について分析して（ステップS27）、その音声波形分析結果を音声採否判定処理6cおよび話者管理処理6aに出力する。

## 【0138】

音声採否判定処理6cでは、上述したように予め設定して音声波形データベース12に保存しておいた音声採否判定基準を読み出し、読み出した音声採否判定基準を用いて、音声波形分析結果が示す高さ（基本周波数）、大きさ（パワー）、速さ（継続時間）のいずれもが音声採否判定基準内に収まっている（採用）か否（不採用）かの判定を行う（ステップS28）。

10

## 【0139】

高さ（基本周波数）、大きさ（パワー）、速さ（継続時間）のいずれもが音声採否判定基準内に収まっている場合（ステップS28でYES）には、音声採用判定処理6cは、録音部7および話者管理処理6a（さらに録音管理処理6dに）に採用信号を出力する。録音部7では、採用信号を入力すると、先に録音しておいた音声波形データを音声波形データベース12に登録する。また、録音管理処理6dでは、採用信号が入力されると、そのときの音声波形データが音声波形データベース12に登録されたのであるから、録音済件数に1を加える。すなわち、（録音済件数+1）を録音済件数として設定する（ステップS29）。

20

## 【0140】

高さ（基本周波数）、大きさ（パワー）、速さ（継続時間）のいずれかが音声採否判定基準内の範囲に収まっていない場合（ステップS28でNO）には、その音声を不採用（録音失敗）として扱われ、音声採否判定処理6cは、不採用信号を話者管理処理6aおよび録音部7に出力する。

## 【0141】

話者管理処理6aは、不採用信号が入力されると、前回不採用の文字列が何番目であったかを示す前回失敗番号と今回の発話の文字列が何番目であるものかを示す（録音済件数+1）とを読み取り、読取った前回失敗番号と（録音済件数+1）との差が予め設定された休息要否判定値未満であるかどうかの判定を行う（ステップS30）。

30

## 【0142】

話者管理処理6aは、（録音済件数+1）と前回失敗番号との差が休息要否判定値以上の場合には、休息は不要で単に録音をやり直せばよい。このときには、前回失敗番号として（録音済件数+1）を設定するとともに、録音のやり直しのために（録音済件数+1）を録音件数として録音管理処理6dに出力する。録音管理処理6dは、（録音済件数+1）番目の文字列を表示装置9に表示させ、もう一度（録音済件数+1）番目の文字列の録音をやりなおす（ステップS34からステップS25に戻る）。

40

## 【0143】

話者管理処理6aは、（録音済件数+1）と前回失敗番号との差が休息要否判定値未満の場合には、頻繁に不採用判定があったのであり、休息が必要であるとして休息指示を生成し、録音管理処理6dに出力する（ステップS31）。録音管理処理6dは出力された休息指示を表示装置9に表示する。話者SPは表示装置9に表示された休息指示を見て、休息する。

## 【0144】

話者管理処理6aの休息指示生成処理62は、休息指示を表示した時点から経過時間の計測を開始して、既定の休息時間が経過するまで待ち状態となる（ステップS32）。経過時間

50

を計測（ステップ S 33）して，休息時間が経過すると（ステップ S 32で Y E S）ステップ S 34へ進み，（録音済件数 + 1）番目の文字列を再度読上げ文字列とする。

【 0 1 4 5 】

以上のように録音済件数が録音全件数に等しくなるまで繰返し録音処理が行われる（ステップ S 23）。

【 0 1 4 6 】

ラベリング部 8 には，原稿作成部 5 で作成された録音原稿と音声波形データベース 12 に保存された音声波形データとが与えられる。ラベリング部 8 は，音声波形データにおいて，その波形に対応する文字列を構成する各音声単位の境界を定め，各音声単位を表わすラベルと，境界を示す時間情報からなるラベル情報を作成する。ラベリング部 8 はまた，作成したラベル情報についてのラベリング・エラー除去（時間情報の修正と時間情報の無効化）を行う。ラベリング部 8 は，ラベル情報をラベル情報データベース 13 に保存する。

10

【 0 1 4 7 】

一例として原稿作成部 5 で作成された録音原稿の中の「さとう ( satoo ) 」という文字列（苗字）を取上げる。音声波形データベース 12 にはこの文字列を話者が発話したときの音声波形データが既に格納されている。音声単位が音素の場合には，上記文字列は，音素を単位としたラベル列 s , a , t , o , o で表わされる。音声単位が音節の場合には，ラベルは sa , to , o となる。ラベリングとは，これらのラベル列の各音声単位と音声波形データとを対応させることであり，音声波形データを，音声単位ごとに区切ることである。音声単位が音素の場合について図 3 を再度参照のこと。

20

【 0 1 4 8 】

図 10 はラベリング部 8 の機能ブロック図である。ラベリング部 8 は，ラベリング処理（手段）8 a と，ラベリングエラー除去処理（手段）8 b とから構成されている。ラベリング処理 8 a には，統計モデル作成処理（手段）81，音声単位境界決定処理（手段）82 およびラベル情報生成処理（手段）83 が含まれている。ラベリングエラー除去処理 8 b には，時間情報エラー修正処理（手段）84，時間情報比較処理（手段）85 およびラベル情報無効化処理（手段）86 が含まれている。

【 0 1 4 9 】

ラベリング処理 8 a の音声単位境界決定処理 82 は，原稿作成部 5 から与えられる録音原稿および音声波形データベース 12 に保存された音声波形データを讀込む。録音原稿は統計モデル作成処理 81 にも与えられる。録音原稿に含まれる一つ一つの文字列（たとえば「sato o」）について次の処理が行なわれる。

30

【 0 1 5 0 】

統計モデル作成処理 81 は，予め用意した統計モデル（音声単位ごとに音響的特徴を統計的にモデル化したもの；たとえば Hidden Markov Model）を利用して，入力された録音原稿中の特定の一つの文字列に対応するラベル列にしたがって，そのラベル列を表わす音声波形に相当する音響的特徴量の系列を作成する。音声単位境界決定処理 82 はこの作成された系列と，上記文字列に対応して実際に録音された音声波形の音響的特徴量の系列とのマッチングをとることによって，実際に録音された（音声波形データベース 12 からの）音声波形において音声単位の境界を抽出する。

40

【 0 1 5 1 】

抽出された音声単位の境界情報（時間情報）は，音声単位を表わすラベルと対にされ，ラベル情報生成処理 83 からラベル情報データベース 13 に与えられる。ラベル情報は，音声単位を表わすラベルとその音声単位の終了時点（時間情報）（音声波形データの開始時点をとる）との対を，文字列の順序（時間の順序）で記述したものである。

【 0 1 5 2 】

なお，ラベリングの詳細については，特開平 10 - 49193 号公報などに開示されている。また，HMM を用いた自動ラベリングのほか，DP マッチングによる自動ラベリングの方式を利用してもよい。

【 0 1 5 3 】

50

ラベリングエラー除去処理 8 b は、生成したラベル情報において、ラベリングエラーの可能性が高い音声単位について、その時間情報（終了時点）を修正したり（時間情報エラー修正）、その音声単位自体をデータベースにおいて無効化するための情報を付与したりする（ラベル情報無効化処理）ものである。すなわちエラー除去の処理内容は、修正規則に基づく時間情報エラー修正と、別個に作成された複数のラベル情報の差異に基づくラベル情報無効化の 2 つに大きく、分けられる。

【 0 1 5 4 】

時間情報エラー修正処理 84 では、予め用意した修正規則により、ラベル情報の時間情報を修正する。

【 0 1 5 5 】

ラベル情報無効化のために、時間情報比較処理 85 では、先の統計モデル（たとえば HMM モデル）を用いて生成したラベル情報（第 1 のラベル情報という）（データベース 13 に格納したもの）と、これとは別の統計モデルを用いて作成した第 2 のラベル情報との差異を比較する。そして、ラベル情報無効化処理 86 において、時間情報の差異が予め設定した閾値を超える場合に、それに対応する第 1 のラベル情報の該当する部分に無効化情報を付与する（無効化情報が付与されたラベル情報は、続く特徴量抽出処理の対象外とされるので、ラベリングエラーが存在したとしても音声単位データベースの品質に悪い影響を与えないようになっている）。

【 0 1 5 6 】

図 18 はラベリング部 8 におけるラベリングエラーの除去処理 8 b の動作を示すフローチャートである。

【 0 1 5 7 】

ラベリングエラー除去処理 8 b はラベリング処理 8 a によって作成され、保存された一文字列（一苗字）についてのラベル情報をラベル情報データベース 13 から読み込む（ステップ S 41）。このラベル情報を第 1 のラベル情報とする。

【 0 1 5 8 】

第 1 のラベル情報に含まれるラベル数をカウントし、このカウント値を変数「全ラベル数」にセットし、ラベル修正規則の数を変数「全規則数」にセットし、変数「処理済ラベル数」を 0 にリセットする（ステップ S 42）。ラベル修正規則については後述する。変数「処理済修正規則数」を 0 にリセットする（ステップ S 44）。ラベル修正規則については、後述する。変数「処理済修正規則数」を 0 にセットする（ステップ S 44）。

【 0 1 5 9 】

（処理済ラベル数 + 1）番目のラベル情報に対して修正規則を順に適用する（ステップ S 46）。修正規則の条件に適合しない場合は、ラベル情報は更新しない。適合する場合は、修正規則の実行部の記述にしたがって、ラベル情報を更新する（ステップ S 47）。

【 0 1 6 0 】

図 19(A) は、ラベル情報の一例を示すものである。ラベル情報は s, a, t, o および o の音声単位（音素）とこれらの音素に対応する境界情報が列記されている。境界情報は、「satoo」の音声波形データの開始時点为零として各音声単位の終了時点の時間情報である。

【 0 1 6 1 】

図 19(B) は、修正規則の例を示すものである。修正規則は各音声単位について設定されている。修正規則は、「if（条件部），then（実行部）」という形式で表現されており、条件部に記述された条件を満足する場合に限って実行部に記述された処理が実行される。

【 0 1 6 2 】

図 19(A) に示すラベル情報について図 19(B) に示す修正規則を具体的に適用してみる。

【 0 1 6 3 】

図 19(A) の第 3 番目のラベル「a」の持続時間は 0.076（秒）（ $0.101 - 0.025 = 0.076$ ）である。図 19(B) のラベル「a」についての修正規則の条件部は「if（持続時間 < 30）」であるから（30 は 0.030 秒の意味）、ラベル「a」の持続時間は条件部を満たさない（ス

10

20

30

40

50

トップ S 46で N O )。したがって、修正規則の実行部は実行されない。

【 0 1 6 4 】

ラベル情報の 5 番目のラベル「 o 」の持続時間は 0.028 ( 0.191 - 0.163 ) であるから、音声単位「 o 」についての修正規則の条件部 ( if ( 持続時間 < 40 ) ) を満たす ( ステップ S 46で Y E S )。したがって、その実行部「修正持続時間 = 持続時間 × 1.5 」が実行される。持続時間の値 = 0.028 であるから、修正持続時間 = 0.042 ( = 0.028 × 1.5 ) となる。5 番目のラベルの「 o 」の終了時点は 0.205 ( = 直前の音声単位の終了時点 0.163 + 0.042 ) と修正される。図 19(C) は、修正した後のラベル情報を示す。

【 0 1 6 5 】

処理済修正規則数に 1 を加えながら ( ステップ S 48 )、( 処理済ラベル数 + 1 ) 番目のラベル情報に対してすべての修正規則を適用する ( ステップ S 45による繰返し)。 10

【 0 1 6 6 】

一つのラベル情報に対してすべての修正規則を適用し終われば、処理済ラベル数に 1 を加え ( ステップ S 49 )、ステップ S 43に戻る。一つの文字列の全ラベルについてステップ S 44 ~ S 49の処理を終えれば ( ステップ S 43で N O )、時間情報のエラー修正処理を終える。

【 0 1 6 7 】

次にラベル情報無効化処理に移る。

【 0 1 6 8 】

第 1 のラベル情報を作成したときに用いた統計モデルとは異なる統計モデルを用いて、第 1 のラベル情報の作成と同じやり方で自動ラベリングを実行し、第 2 のラベル情報を作成する ( ステップ S 50 )。作成された第 2 のラベル情報の例が図 20(A) に示されている。 20

【 0 1 6 9 】

処理済ラベル数を 0 に戻し、ラベル無効化の閾値を設定する ( ステップ S 51 )。ラベル無効化閾値の例が図 20(C) に示されている。

【 0 1 7 0 】

修正された第 1 のラベル情報 ( 図 19(C) ) と第 2 のラベル情報 ( 図 20(A) ) とにおいて、対応するラベルの時間情報の差をそれぞれ算出し ( ステップ S 53 )、この差がラベル無効化閾値を超えているかどうかを判定する ( ステップ S 54 )。各ラベルについての差の一例が図 20(B) に示されている。これらの差のうち閾値を超えているものがあれば ( ステップ S 54で Y E S )、対応する第 1 のラベル情報に無効化情報を付与する ( ステップ S 55 )。例えば、図 20(B) において、2 番目のラベル「 a 」の時間情報の差は、0.014 ( s ) であり、ラベル無効化閾値である 0.050 ( s ) の範囲に収まっているので、無効化情報を付与する必要はない。これに対して 5 番目のラベル「 o 」の時間情報の差は、0.051 ( s ) であり、ラベル無効化閾値を超えているため、5 番目の「 o 」のラベルに無効化情報を付与する。その直後のラベル ( 6 番目のラベル「 o 」 ) のラベルにも自動的に無効化情報を付与する。図 20(D) では、第 5 番目と第 6 番目のラベル「 o 」に無効化情報 × が付けられている。処理済ラベル数に 1 を加えながらすべてのラベルについて上記の処理を繰返す ( ステップ S 56, S 52 )。無効化処理を終えた第 1 のラベル情報はデータベース 13 に再び格納される。 30 40

【 0 1 7 1 】

図 21 は、音声情報データベース 15 に含まれるインデックス情報 ( 図 21(A) ) と、これに対応する音声波形データ ( 図 21(B) ) の一例を示したものである。

【 0 1 7 2 】

特徴量抽出部 14 は、ラベル情報データベース 13 に保存されたラベル情報を読み出し、対応する音声波形データを音声波形データベース 12 から読み出す。特徴量抽出部 14 は、読み出した対応するラベル情報と音声波形について、音声単位ごとに、特徴量 ( 長さ、高さ、大きさなど ) を算出して、算出した特徴量をラベル情報とともに列記してインデックス情報を作成する。このとき上述した無効化情報が付与された音声単位については特徴量の算出は行わない。さらに特徴量抽出部 14 はインデックス情報と音声波形データとを対にして音声情報 50

データとして、音声情報データベース15に保存する。

【0173】

(3) 第2実施例

図22は、音声情報データ作成システムの第2実施例の全体構成を示す機能ブロック図である。この図において、図7に示すものと同一物には同一符号を付し重複説明を避ける。録音原稿を話者SPが声を出して読むにあたって適切な読み方を話者SPに示すための標準音声を作成する機能を録音管理部6Aが持つ。標準音声はスピーカ18から出力される。原稿作成部5Aは、元原稿を追加したときに、既に作成されている録音原稿に追加すべき追加録音原稿を作成する機能を持つ。この追加録音原稿は最小限で足りる。追加録音原稿の作成のために音声情報データベース15から音声情報(インデックス情報)が原稿作成部5Aに与えられる。ラベリング部8Aは第1の実施例とは異なり、作成したラベル情報(特に時間情報)について、ラベル情報の統計的分析結果に基づいてエラー除去を行う機能を持つ。

10

【0174】

図23は原稿作成部5Aの機能的構成を示すブロック図である。図24は原稿作成部5Aの追加録音原稿を作成する動作を示すフローチャートである。以下に、追加録音原稿を作成する処理について説明する。録音原稿作成処理は第1実施例において説明した通りであり、追加録音原稿作成処理はこれに付加される機能であると理解されたい。

【0175】

以下の説明では、第1実施例において既に作成された苗字についての追加録音原稿の存在を前提とする。

20

【0176】

音声情報データベース15には苗字についての録音原稿を話者SPが読上げて、これを録音して得られる音声情報が既に格納されているものとする。音声情報データベース分析処理(手段)54は、データベース15から苗字についての音声情報中のインデックス情報を読み出し、このインデックス情報を分析してインデックス情報に含まれる音声単位のリストを、品質レベルごとに作成する(図24、ステップS61)。音声情報データベースの分析結果の一例が図25(A)に示されている。これは図13(B)に示す元原稿分析結果と全く同じである(音声単位のリストにおける配列順序が異なっているが)。

【0177】

30

仕様入力部4において、現在の音声情報データベース15を前提として、新たに追加的に合成により得たい文字列(単語、句、節、文などを含む)を列挙した原稿(追加元原稿という)の入力を操作者OPから受け付ける(ステップS62)。入力が終了するまでは、待ち状態にある(ステップS63)。追加元原稿に対応するテキストファイル名だけを入力装置42において入力させて、追加元原稿の内容は、そのファイルをFDドライブ41により読み込むようにしてもよい。もちろん追加元原稿をキーボードから入力してもよいし、原稿データベース11に格納されているものを用いてもよい。

【0178】

追加元原稿の一例が図25(B)に示されている。この追加元原稿は地名リストである。

【0179】

40

追加元原稿が元原稿設定処理51Aに設定されると、元原稿分析処理52Aは、追加元原稿に含まれるすべての文字列について、それらを品質レベル別に、ラベル(音声単位)に分解し、それらの出現回数を計数して、音声単位リストを作成する(ステップS64)。これが追加元原稿分析結果であり、図25(B)に示す追加元原稿について、具体例が図25(C)に示されている。

【0180】

分析結果比較処理55Aは、元原稿分析処理52Aによる追加元原稿分析結果と音声情報データベース分析処理54による音声情報データベース分析結果とを比較し、追加元原稿分析結果(図25(B))に存在するが、音声情報データベース分析結果(図25(A))には存在しない音声単位を、品質レベル別に抽出する。この差分抽出結果の一例が図25(D)に示されて

50

いる。

【0181】

文字列選択処理（手段）53Aは、差分抽出結果に含まれている音声単位について、その音声単位を含む文字列を、録音原稿に追加していく、という処理を品質レベル別に品質レベルの低い方から高い方に向けて順に行う。追加録音原稿が、その品質レベルの全音声単位を網羅した時点で、その品質レベルの処理を終え、次の品質レベルの処理に移る。図25(D)に示す例では、品質レベル1には差分として抽出された音声単位がないので、品質レベル2から処理を行うことになる。品質レベル2における処理では「きょうと」が追加され、品質レベル3における処理ではさらに「なら」が追加され、最終的には、追加録音原稿には、「きょうと」と「なら」の2つの文字列が追加される。音声情報データベースにこの2つの文字列の音声を追加するだけで、図25(A)の地名リスト中のすべての地名を、高い品質で合成することができるようになる。

10

【0182】

このようにして、追加原稿の分析結果とインデックス情報分析結果とを比較して、追加元原稿にあってインデックス情報にない音声単位（不足音声単位）を抽出し、不足している音声単位を含む文字列を録音原稿に追加していくので、最初から録音原稿を作り直す必要はない。

【0183】

図24において、分析結果比較処理55と文字列選択処理53Aの動作の流れは次の通りである。

20

【0184】

音声情報データベース分析結果と追加元原稿分析結果を参照し、追加元原稿に存在して、音声情報データベースには存在しないような音声単位をすべて列挙し、音声単位リストとする。また、追加元原稿に含まれるすべての文字列を、文字列リストに加える（ステップS65）。

【0185】

音声単位リストに音声単位が残っている場合（ステップS66でYES）、音声単位リストから出現回数が最小の音声単位1つを選択し、さらに、文字列リストからその音声単位を含む文字列を1つだけ選択して、追加録音原稿にその文字列を追加する（ステップS67）。

30

【0186】

追加録音原稿に追加した文字列に含まれる音声単位のうち、音声単位リストに残っているものをすべて音声単位リストから削除する（ステップS68）。また、追加録音原稿に追加した文字列を文字列リストから削除する（ステップS69）。

【0187】

音声単位リストが空になるまでステップS67～S69を繰り返す。これにより、追加録音原稿作成が終了する。

【0188】

この追加録音原稿作成処理においても、データベース容量やデータベース作成時間の要求がある場合は、この要求による制限が考慮されるのはいうまでもない。

40

【0189】

図26は録音管理部6Aの構成を示すブロック図である。

【0190】

録音管理部6Aは、上述した第1実施例の録音管理部6に音声合成処理（手段）6eがさらに設けられたものである。

【0191】

音声合成処理6eは、原稿作成部3Aから録音管理処理6dを介して読込んだ録音原稿中の文字列を音で表わす合成音声を作成する。すなわち、音声合成処理6eは録音原稿の各文字列について、正しい読み上げ方（アクセント位置、間の取り方、抑揚などの点で）、または（話者管理手段6cが保持する録音音声分析結果の履歴情報に基づいて）その話者

50

に適切な声の大きさ，高さ，速さで，録音原稿の文字列を読み上げる合成音声を作成する。音声合成手段 6 e で作成された合成音声（予め用意した録音音声でもよい）は標準音声としてスピーカなどの音声出力装置 18 から出力される。これにより，話者 S P は，発話すべき文字列の合成音声を聞き，発話すべき音声の高さ，大きさ，速さを参考にすることができるため，文字列を不適切に読むことを防ぎ，録音音声（音声情報データ）の質を向上させることができる。

【 0 1 9 2 】

図 27 は，録音管理部 6 A による録音管理処理を示すフローチャートである。図 15 に示すものと同一処理には同一符号を付し重複説明を避ける。また，図 16 はそのまま適用することができる。

10

【 0 1 9 3 】

音声合成手段 6 e は，録音管理処理 6 d から入力された（録音済件数 + 1）番目の文字列に対する適切な声の高さ，大きさ，速さ，抑揚等の目標値，またはこれまでの録音音声の分析結果に基づき，韻律的特徴のパラメータを設定する（ステップ S 36）。音声合成手段 6 e は，設定したパラメータを用いて（録音件数 + 1）番目の読上げ文字列の合成音声を作成し，作成した合成音声を標準音声として音声出力装置（スピーカ）18 に出力する（ステップ S 37）。したがって，表示画面に文字列が表示される（ステップ S 25）だけでなく，その標準音声も出力される。

【 0 1 9 4 】

図 28 はラベリング部 8 A の機能的構成を示すブロック図である。図 10 と比較すると，ラベリングエラー除去処理 8 b に代えて，ラベリングエラー除去処理（手段）8 c およびラベル情報統計分析処理（手段）8 d が設けられている。ラベリングエラー除去処理 8 c は，ラベル情報信頼性確認処理（手段）87 およびラベル情報無効化処理（手段）86 を含む。ラベル情報統計分析処理 8 d は，信頼区間算出処理（手段）88 および統計分析処理（手段）89 を含む。

20

【 0 1 9 5 】

ラベリング情報統計分析処理 8 d は，既存のラベル情報（ラベル情報データベース 13 内のラベル情報）を統計的に分析して，音声単位ごとに継続時間の平均値と標準偏差から信頼区間（継続時間に関する信頼区間）を算出し，信頼区間情報を作成する。分析対象となる既存のラベル情報は，話者により音声の特徴が異なり，継続時間の信頼区間が変わることが多いため，これからラベリングエラー除去の処理をしようとするラベル情報と同じ話者のラベル情報を使用することが望ましい。

30

【 0 1 9 6 】

ラベリングエラー除去処理 8 c は，ラベル情報統計分析処理 8 d で得られた各音声単位の信頼区間情報を参照して，エラー除去対象のラベル情報に含まれる各音声単位の継続時間が対応するラベルの信頼区間内に収まっているか否かをチェックする。ラベリングエラー除去処理 8 c は，信頼区間に収まっていないラベル情報に無効化情報を付与する，さらに信頼区間内に収まるようにラベル時間情報を修正してもよい。ラベリングエラー除去処理 8 c は，ラベリング処理 8 a により生成されたラベル情報に含まれる各音声単位について，その音声単位の継続時間が，ラベル情報統計分析処理 8 d により算出された，その音声単位に対応する継続時間の信頼区間の範囲外にある場合，その箇所を（ラベリングの信頼性が低い，すなわち，ラベリングエラーの可能性が高いと判定して），無効化する。これにより，統計的に信頼性が低いと判断されたラベル情報を自動的に無効化でき，結果として，ラベリング結果の品質を高めることができる。

40

【 0 1 9 7 】

図 29 は，ラベリング部 8 A のラベル情報統計分析処理 8 d およびラベリングエラー除去処理 8 c によるラベリングエラー除去処理の手順を示すフローチャートである。

【 0 1 9 8 】

ラベル情報統計分析処理 8 d の統計分析処理 89 はラベリング処理 8 a によって作成され，ラベル情報データベース 13 に保存されたラベル情報，好ましくは同じ話者 S P によって録

50



音された音声波形から得られた一群のラベル情報を読み込む（ステップ S 71）。

【 0 1 9 9 】

統計分析処理 89は、音声単位別に継続時間の平均値と標準偏差を算出し、ラベル情報の中にその音声単位が出現する個数を計数する（ラベル情報の統計分析）（ステップ S 72）。

【 0 2 0 0 】

図30(A) は統計分析処理89に読み込まれたラベル情報の一例を示すものである。図30(B) は統計分析処理89による統計分析の結果の一例を示すものである。

【 0 2 0 1 】

信頼区間算出処理 88は、統計分析処理 89による統計分析結果に基づき音声単位ごとの継続時間の信頼区間を以下の算出式により算出する（ステップ S 73）。

【 0 2 0 2 】

信頼区間 = 平均値  $\pm Z [ (標準偏差)^2 / (出現回数) ]^{1/2}$  (式 1)

【 0 2 0 3 】

ここで、Z は正規分布に基づく定数である。

【 0 2 0 4 】

図30(C) は上記の算出式より算出した音声単位ごとの継続時間の信頼区間の一例を示している。

【 0 2 0 5 】

このようにして得られた信頼区間に関するデータはラベル情報信頼性確認処理 87に与えられる。ラベル情報信頼性確認処理 87はまた、統計分析処理 89が取得したものと同一ラベル情報（これをエラー除去対象ラベル情報という）をラベル情報データベース 13から読み込む。

【 0 2 0 6 】

ラベル情報信頼性確認処理 87は、エラー除去対象ラベル情報に含まれるラベル数をカウントし、変数「全ラベル数」に設定する。また、「処理済ラベル数」を 0 に設定する（ステップ S 74）。

【 0 2 0 7 】

（処理済ラベル数 + 1）番目のラベルに対応する音声単位の継続時間を算出する（継続時間は、その音声単位の終点を示す時間情報と、直前の音声単位の終点を示す時間情報との差で求められる）（ステップ S 76）。

【 0 2 0 8 】

（処理済ラベル数 + 1）番目のラベルに対応する音声単位の継続時間が、その音声単位の信頼区間の範囲内におさまっていない場合は、ラベル情報無効化処理 86は、（処理済ラベル数 + 1）番目のラベルに無効化情報を付与する（ステップ S 78）。図30(D) は無効化情報付与後のラベル情報の一例を示すものである。音声単位「o」の継続時間の信頼区間は、図30(C)によると、46.8 ~ 115.2 (ms) である。図30(D)において第5番目および第6番目の音声単位（ラベル）「o」の継続時間はそれぞれ0.191 (s) および0.312 (s) であり、信頼区間の範囲内に入っていない。したがってラベル「o」は2つとも無効化情報（×印で示す）が付与される。他のラベル s, a, t の継続時間は対応する信頼区間の範囲内にあるので無効化情報は付与されない。

【 0 2 0 9 】

処理済ラベル数の値に 1 を加え、ステップ S 75を経てステップ S 76に戻り（ステップ S 79）、ステップ S 76 ~ S 78の処理を、処理済ラベル数が全ラベル数に等しくなるまで繰り返す（ステップ S 75）。

【 0 2 1 0 】

以上のようにして、ラベリングエラー除去処理が終了すると、処理後のラベル情報は再びラベル情報データベース 13に格納される。

【 図面の簡単な説明 】

【 図 1 】 音声波形における音素、音節および単語の区切りを付け、音声波形と対応付けて示す。

10

20

30

40

50

【図 2】音韻環境を考慮して音声単位を音声波形と対応付けて示す。  
 【図 3】音声波形とラベル情報との関係を示す。  
 【図 4】波形情報を用いて波形接続型音声合成を行う様子を示す。  
 【図 5】波形接続型音声合成処理の流れを示すものである。  
 【図 6】音声情報データベース作成システムのハードウェア構成を示すブロック図である。

【図 7】第 1 実施例における音声単位データ作成システムの全体構成を示すブロック図である。

【図 8】原稿作成部の機能的構成を示すブロック図である。

【図 9】録音管理部の機能的構成を示すブロック図である。

10

【図 10】ラベリング部の機能的構成を示すブロック図である。

【図 11】原稿作成部による録音原稿作成処理を示すフローチャートである。

【図 12】仕様入力表示画面を示す。

【図 13】(A) は、元原稿の一例を示す。(B) は、元原稿分析結果の一例を示す。

【図 14】(A) は、レベル 1 処理後の録音原稿の一例を示す。(B) は、レベル 2 処理後の録音原稿の一例を示す。(C) は、レベル 3 処理後の録音原稿の一例を示す。

【図 15】録音管理部による録音処理を示すフローチャートである。

【図 16】録音管理部による録音処理を示すフローチャートである。

【図 17】録音画面を示す。

【図 18】ラベリング部によるラベリングエラー除去処理を示すフローチャートである。

20

【図 19】(A) は、第 1 のラベル情報を示す。(B) は、修正規則を示す。(C) は、修正後の第 1 のラベル情報を示す。

【図 20】(A) は、第 2 のラベル情報を示す。(B) は、ラベルの差異情報を示す。(C) は、無効化閾値を示す。(D) は、無効化情報付与後の第 1 のラベル情報を示す。

【図 21】(A) は、インデックス情報を示す。(B) は、音声波形データを示す。

【図 22】第 2 実施例における音声単位データ作成システムの全体構成を示すブロック図である。

【図 23】第 2 実施例における原稿作成部の機能的構成を示すブロック図である。

【図 24】第 2 実施例における原稿作成部による追加録音原稿作成処理を示すフローチャートである。

30

【図 25】(A) は、音声情報データベース分析結果の一例を示す。(B) は、追加元原稿の一例を示す。(C) は、追加元原稿分析結果の一例を示す。(D) は、差分抽出結果の一例を示す。(E) は、品質レベル 2 処理後の追加元原稿の一例を示す。(F) は、品質レベル 3 処理後の追加元原稿の一例を示す。

【図 26】第 2 実施例における録音管理部の機能的構成を示すブロック図である。

【図 27】第 2 実施例における録音管理部による録音処理を示すフローチャートである。

【図 28】第 2 実施例におけるラベリング部の機能的構成を示すブロック図である。

【図 29】第 2 実施例におけるラベリング部によるラベリングエラー除去処理を示すフローチャートである。

【図 30】(A) は、ラベル情報の一例を示す。(B) は、統計分析結果の一例を示す。(C) は、信頼区間情報の一例を示す。(D) は、無効化情報付与後のラベル情報の一例を示す。

40

【符号の説明】

4 仕様入力部

5, 5 A 原稿作成部

5 a 原稿作成処理

5 b 音声情報データベース分析処理

6, 6 A 録音管理部

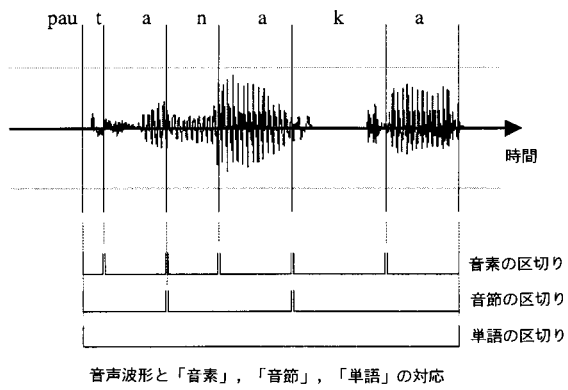
6 a 話者管理処理

6 b 音声分析処理

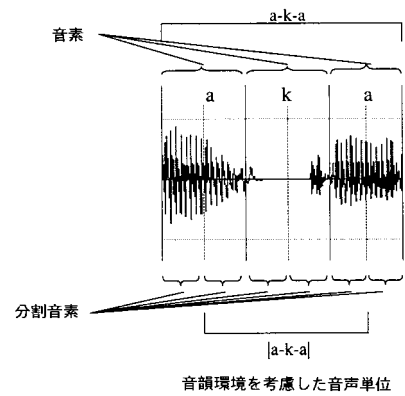
50

- 6 c 音声採否判定処理
- 6 d 録音管理処理
- 6 e 音声合成処理
- 7 録音部
- 8 , 8 A ラベリング部
- 8 a ラベリング処理
- 8 b , 8 c ラベリングエラー除去処理
- 8 d ラベル情報統計分析処理
- 9 表示装置
- 10 音声入力装置
- 11 原稿データベース
- 12 音声波形データベース
- 13 ラベル情報データベース
- 14 特徴量抽出部
- 15 音声情報データベース
- 16 出力装置
- 17 記録媒体

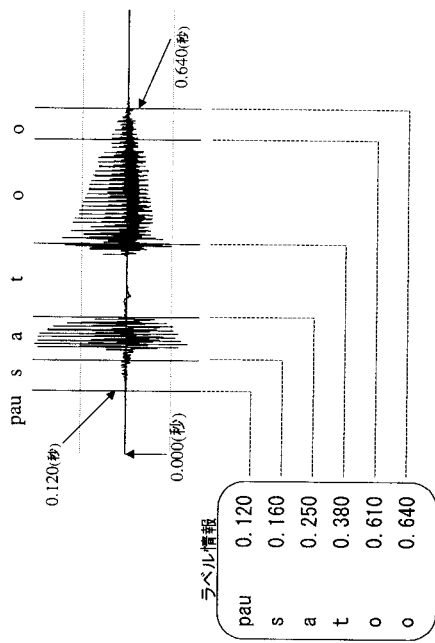
【 図 1 】



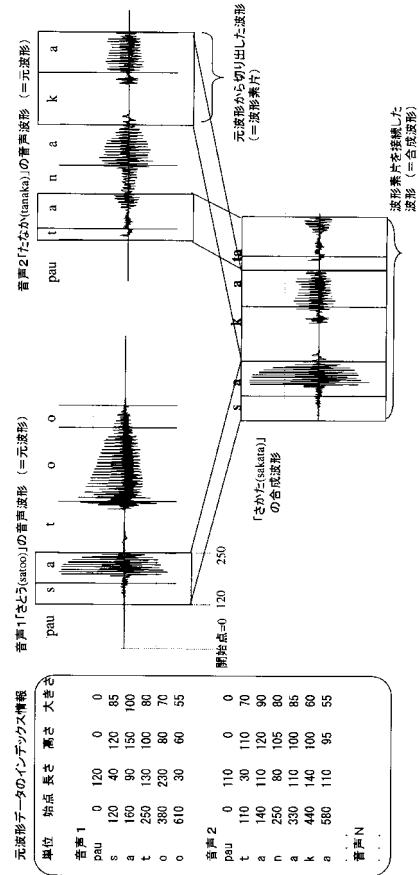
【 図 2 】



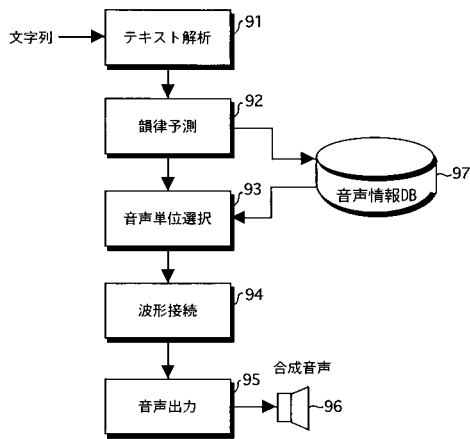
【図3】



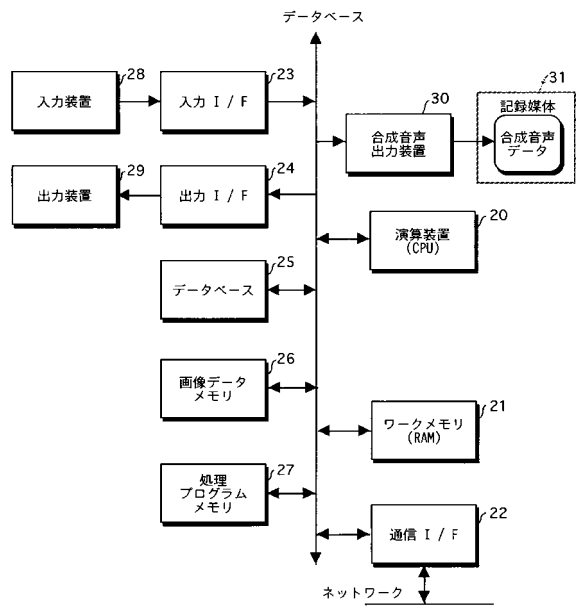
【図4】



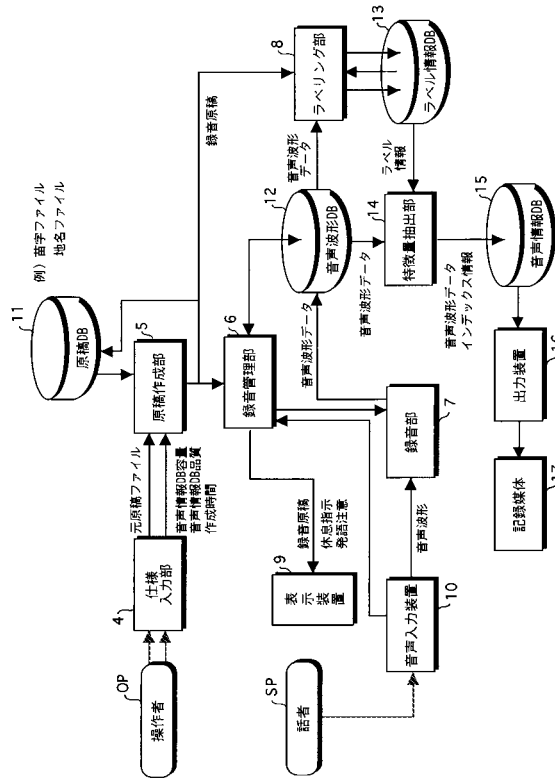
【図5】



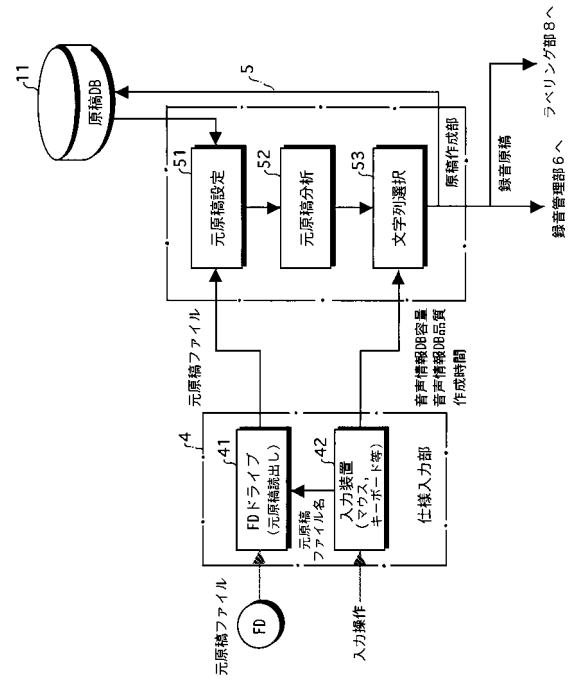
【図6】



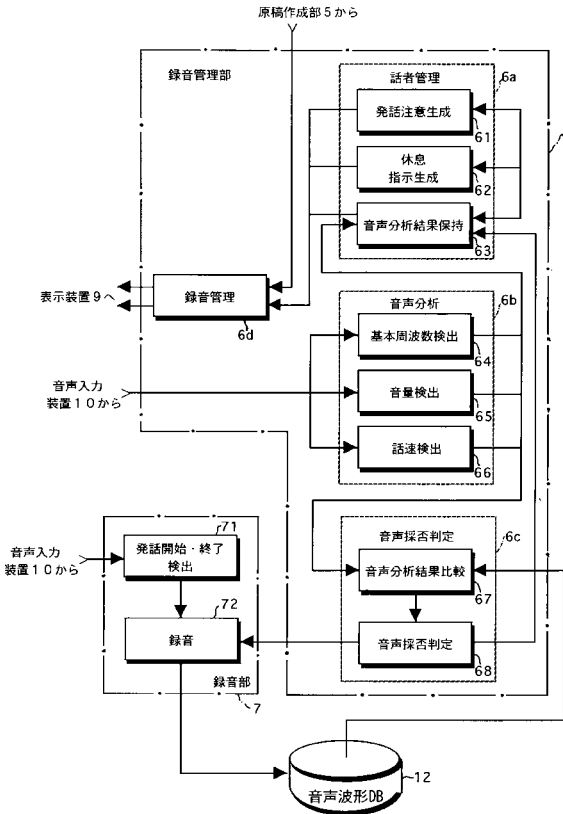
【図7】



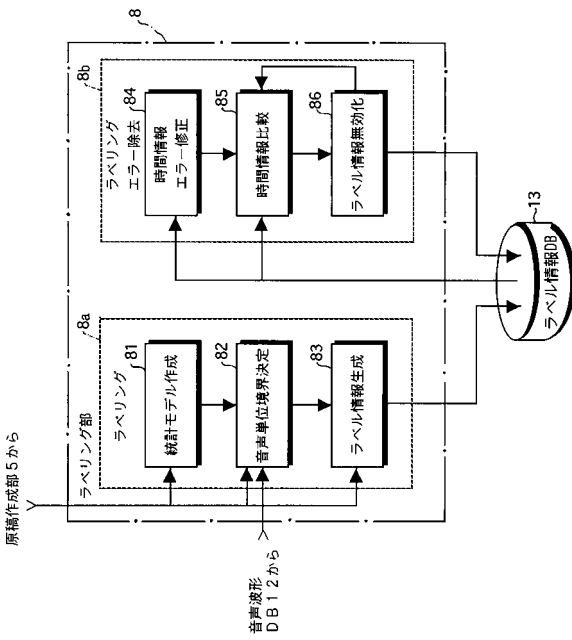
【図8】



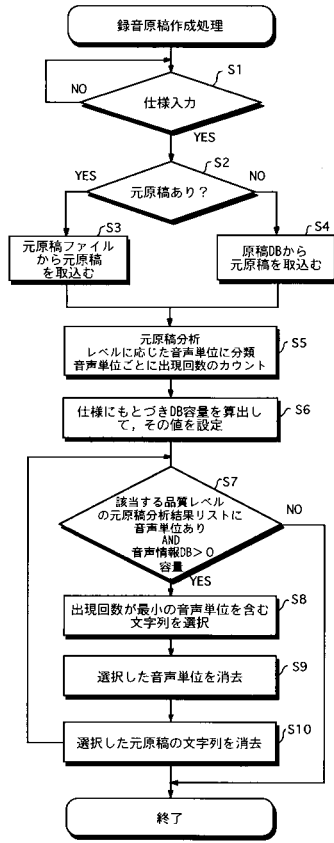
【図9】



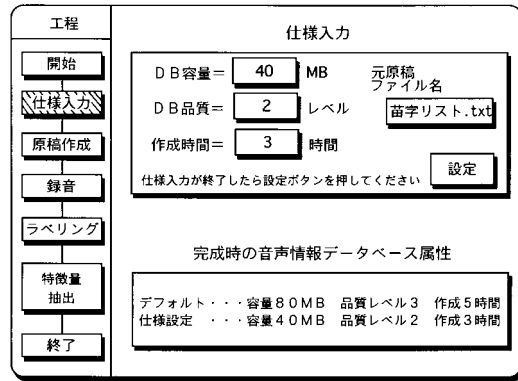
【図10】



【 図 1 1 】



【 図 1 2 】



【 図 1 3 】

(B)

品質レベル1	品質レベル2	品質レベル3
b e w y z n r sh s u k i m t o a	1 1 1 1 1 1 2 2 3 3 3 4 4 5 6 6 9 14 20	be ka ki ma su wa ha ra ya zu mi na ka to

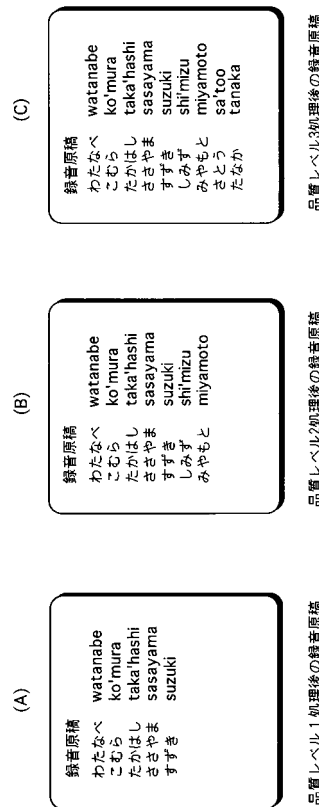
(A)

さとう すずき たかはし たなか わたなべ いとう ささやま しみず みやもと なかもと はしimoto おおた こむら むらかみ

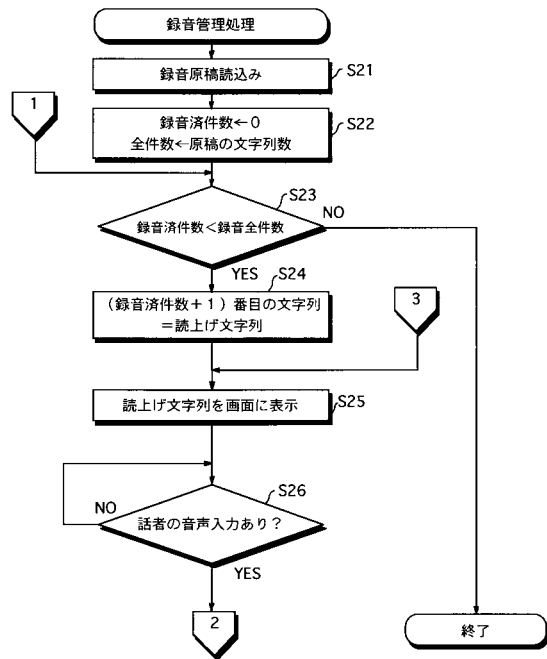
元原稿

元原稿分析結果

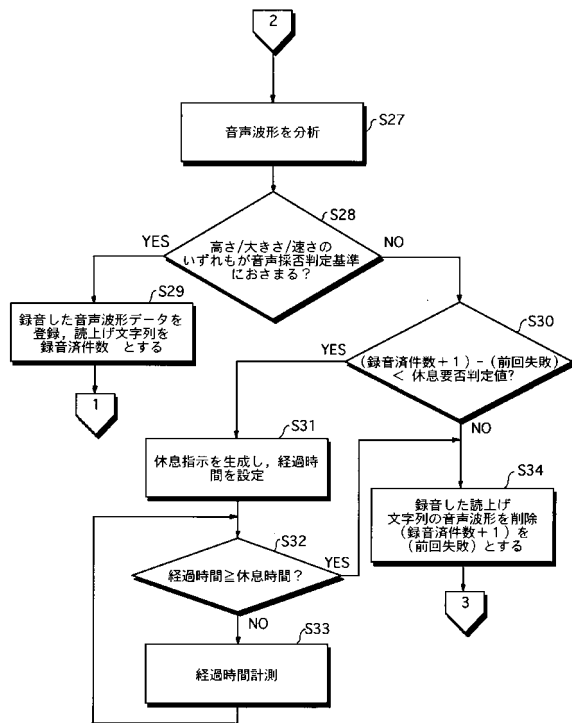
【 図 1 4 】



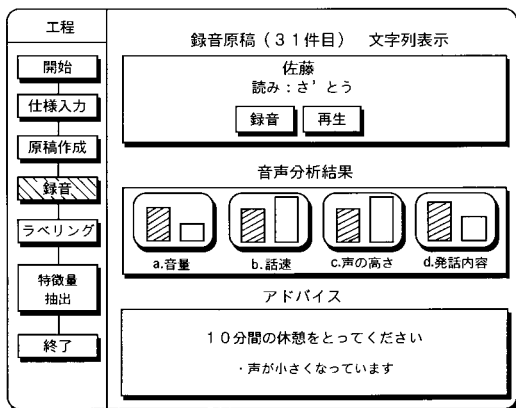
【 図 1 5 】



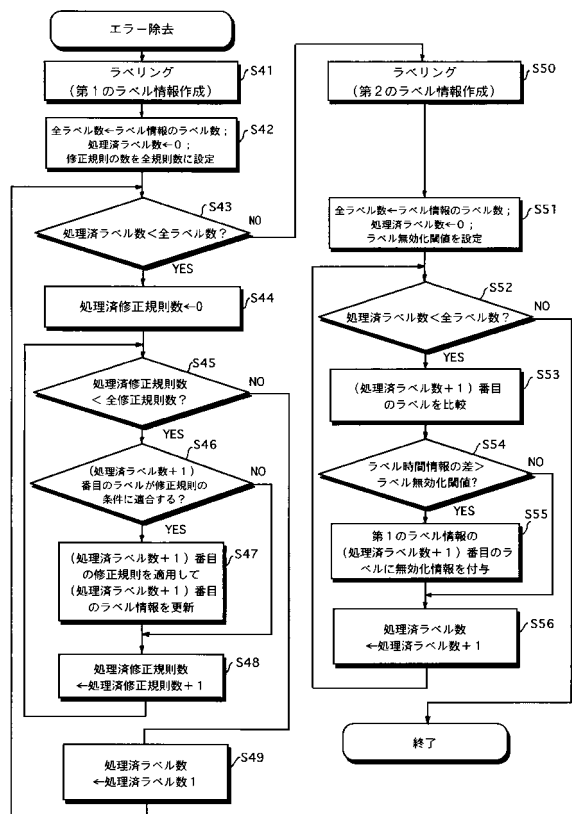
【 図 1 6 】



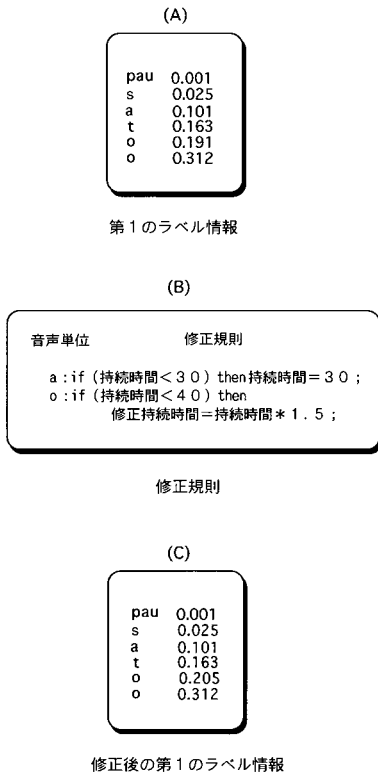
【 図 1 7 】



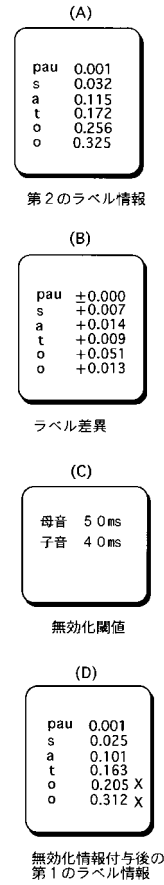
【 図 1 8 】



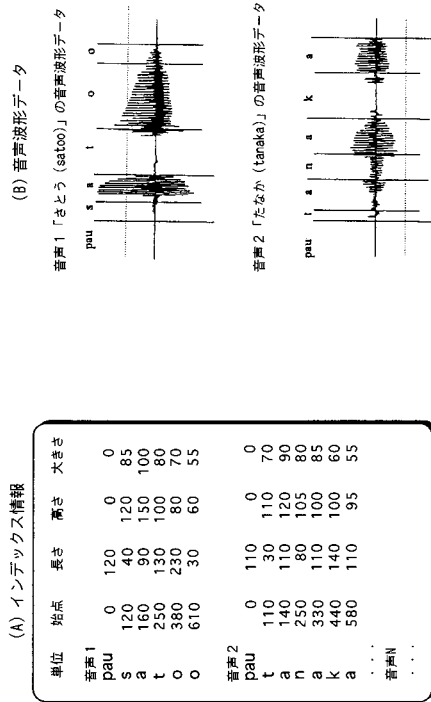
【図19】



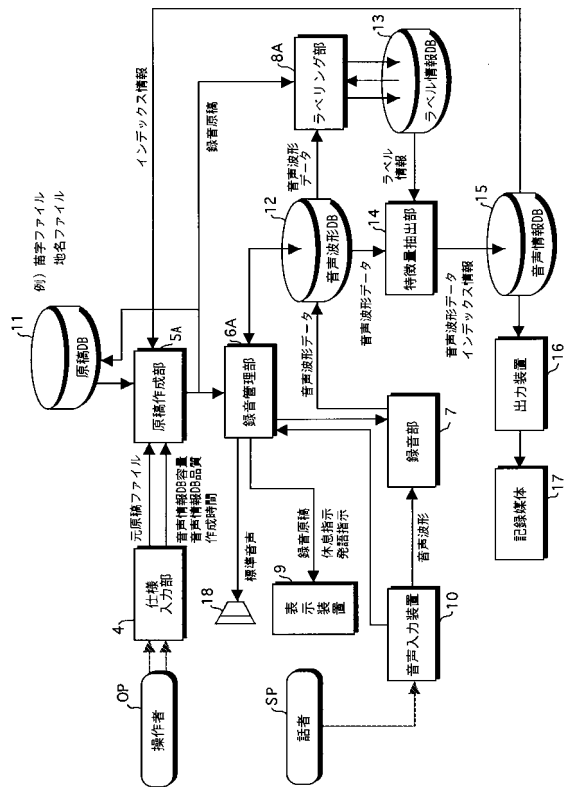
【図20】



【図21】

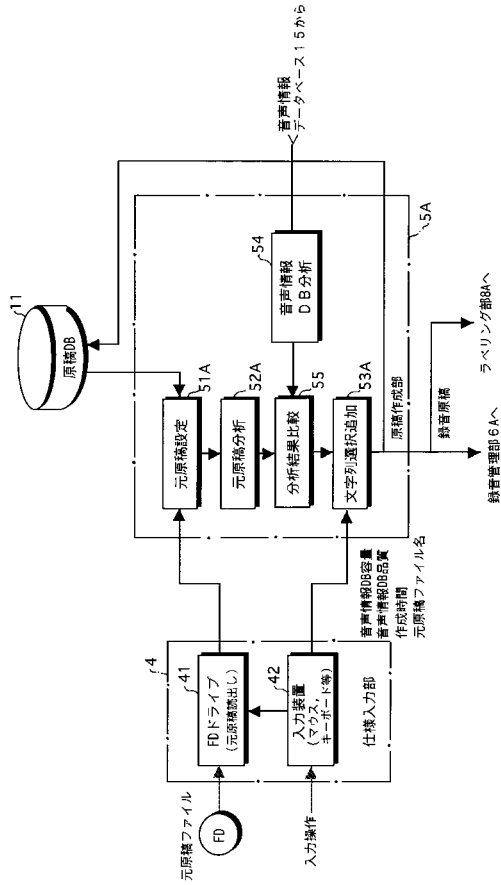


【図22】

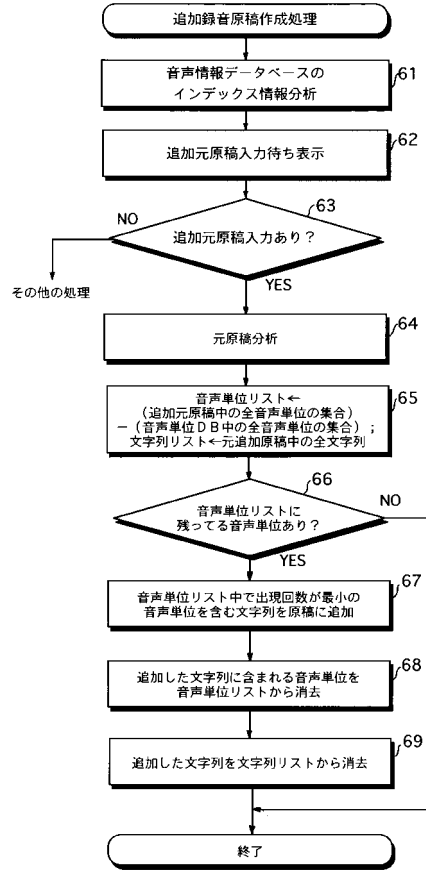




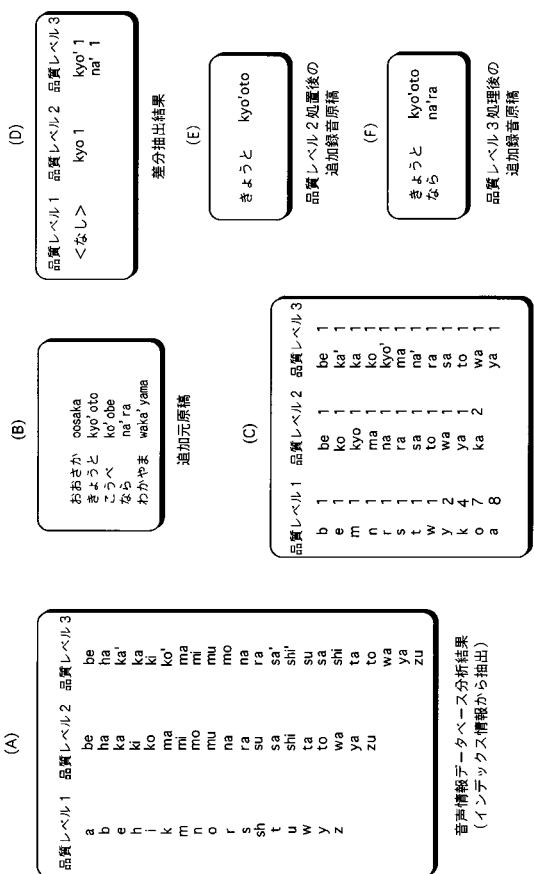
【 図 2 3 】



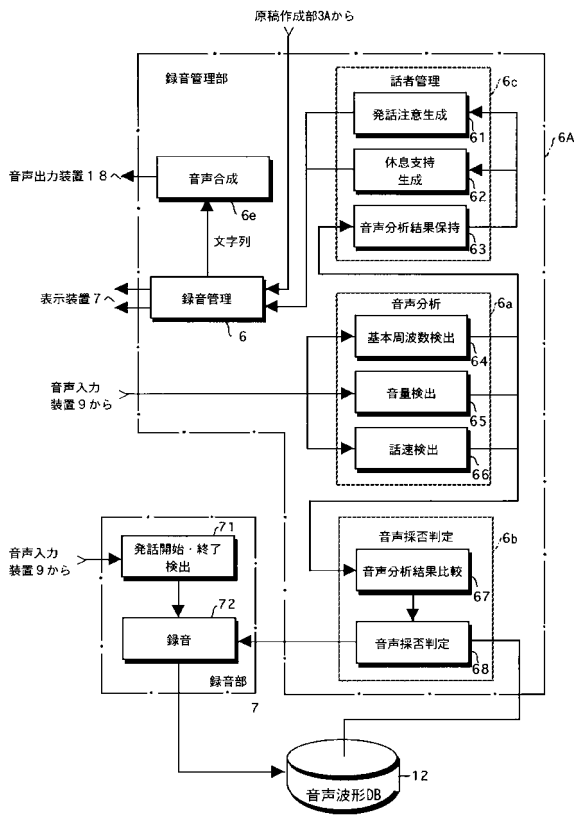
【 図 2 4 】



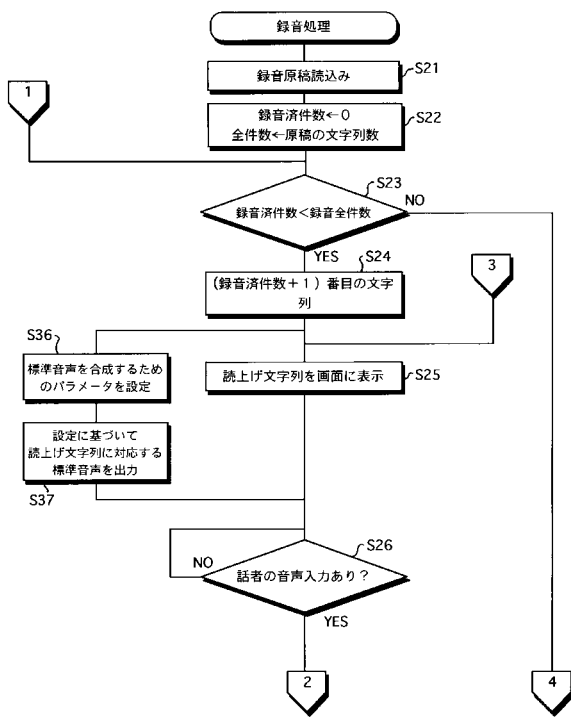
【 図 2 5 】



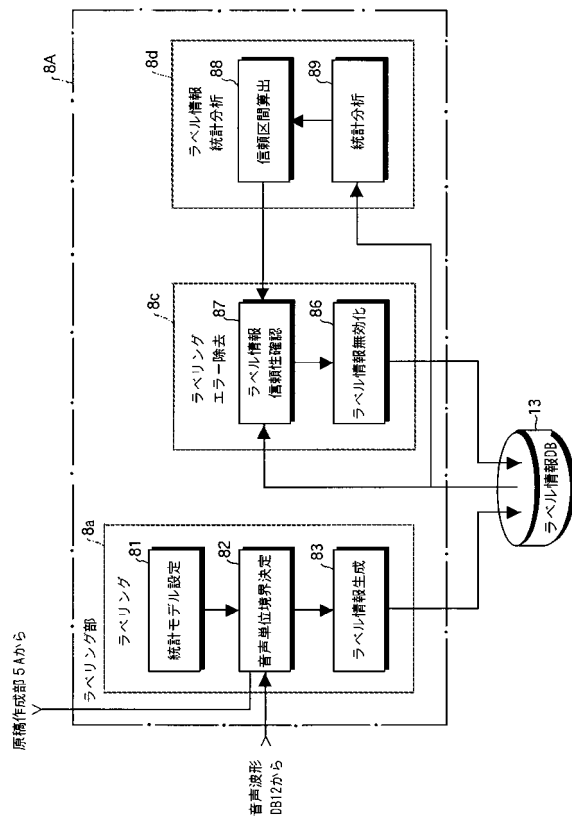
【 図 2 6 】



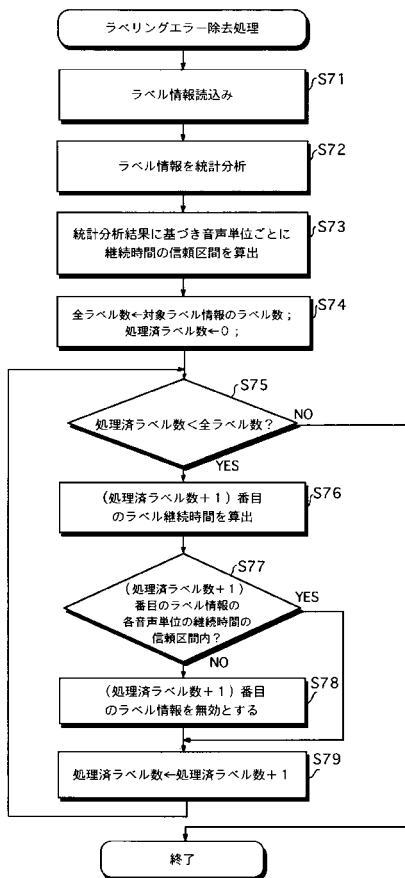
【 図 27 】



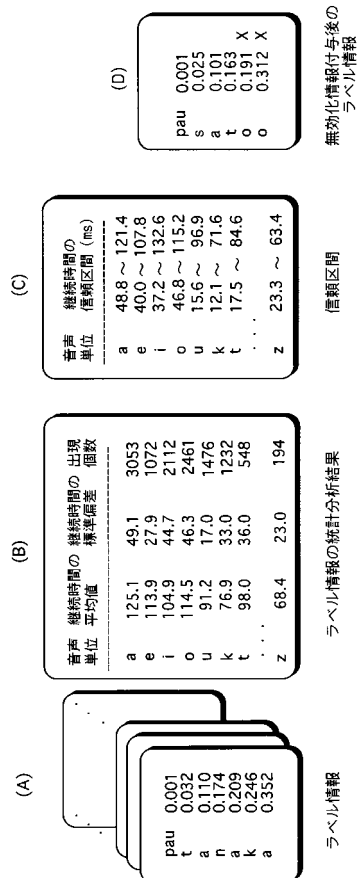
【 図 28 】



【 図 29 】



【 図 30 】



---

フロントページの続き

- (72)発明者 堤 ゆみ  
京都市下京区塩小路通堀川東入南不動堂町801番地 オムロン株式会社内
- (72)発明者 大谷 賢  
京都市下京区塩小路通堀川東入南不動堂町801番地 オムロン株式会社内
- (72)発明者 糺谷 和人  
京都市下京区塩小路通堀川東入南不動堂町801番地 オムロン株式会社内

審査官 山下 剛史

- (56)参考文献 特開2001-296878(JP,A)  
特開平8-63189(JP,A)  
特開2000-250573(JP,A)  
特開平7-44187(JP,A)  
特開平8-263520(JP,A)  
特開平8-160991(JP,A)

- (58)調査した分野(Int.Cl., DB名)  
G10L 13/06,15/04  
G06F 17/30