(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4580297号
(P4580297)
(45)発行日　平成22年11月10日(2010.11.10)　(24)登録日　平成22年9月3日(2010.9.3)

(51)Int.Cl.　　F I
G10L 21/04 (2006.01)　　G10L 21/04 110Z
G10L 11/00 (2006.01)　　G10L 11/00 402L

請求項の数 8　(全 24 頁)

(21)出願番号　特願2005-204211(P2005-204211)
(22)出願日　平成17年7月13日(2005.7.13)
(65)公開番号　特開2007-25039(P2007-25039A)
(43)公開日　平成19年2月1日(2007.2.1)
審査請求日　平成20年5月29日(2008.5.29)

(73)特許権者　000005821
パナソニック株式会社
大阪府門真市大字門真１００６番地
(74)代理人　100098291
弁理士　小笠原　史朗
(72)発明者　三崎　正之
大阪府門真市大字門真１００６番地　松下電器産業株式会社内
(72)発明者　正木　芽衣子
大阪府門真市大字門真１００６番地　松下電器産業株式会社内
(72)発明者　河村　岳
大阪府門真市大字門真１００６番地　松下電器産業株式会社内

最終頁に続く

(54)【発明の名称】音声再生装置、音声録音再生装置、およびそれらの方法、記録媒体、集積回路

(57)【特許請求の範囲】
【請求項１】
入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するための音声再生装置であって、
前記音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する判別部と、
前記音声区間および前記非音声区間に基づき、算出用フレーム長に対して当該音声区間が含まれる比率を示す音声含有率と、前記算出用フレーム長の音声含有率の平均値および標準偏差である統計値を少なくとも算出する音声情報算出部と、
前記音声速度変換処理の比率が１以上の速度比を基準値として予め設定し、前記統計値および前記算出用フレーム長の音声含有率に応じて、前記基準値から前記音声区間の速度比を算出する速度比算出部とを備え、
前記音声情報算出部は、それぞれ時間長が異なる前記算出用フレーム長を複数設定してそれぞれ前記音声含有率および前記統計値を算出し、
前記速度比算出部は、前記算出用フレーム長における音声含有率の前記平均値に対する差分値および前記標準偏差によって前記算出用フレーム長毎に得られる値を用いて算出される係数を前記速度比の基準値に乗ずることで、前記算出用フレーム長の音声含有率が相対的に高いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より小さな値に変更し、前記算出用フレーム長の音声含有率が相対的に低いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より大きな値に変更し、音声含有

率の時間的な変動に対して適応的に前記音声区間の速度比を調整する速度制御を行う音声再生装置。

【請求項２】

前記速度比算出部は、ユーザの操作に応じて前記短縮された再生時間を設定し、音声区間における前記変更した速度比に基づいて、前記音声信号の再生時間が前記設定された再生時間となるように前記非音声区間の速度比を算出することを特徴とする、請求項１に記載の音声再生装置。

【請求項３】

前記速度比算出部は、前記設定された再生時間内において前記非音声区間の速度比を一定に算出することを特徴とする、請求項２に記載の音声再生装置。

【請求項４】

前記入力される音声信号のうち、少なくとも前記算出用フレーム長分の音声信号を含むように当該音声信号を順次更新しながら記録するバッファと、

前記バッファに記録された音声信号に対して音声速度変換処理を行って出力する速度変換部とを、さらに備え、

前記判別部は、前記バッファに記録された前記算出用フレーム長の音声信号に対して前記音声区間と前記非音声区間とを判別し、

前記音声情報算出部は、これまでに算出した統計値を単位時間毎に順次更新し、

前記速度比算出部は、前記単位時間ごとに更新される前記統計値および当該更新時の前記算出用フレーム長に設定された音声含有率に応じて前記音声区間の速度比を算出し、

前記速度変換部は、前記バッファで順次更新される音声信号に対して、前記単位時間ごとに算出された前記音声区間の速度比を用いて順次音声速度変換処理を行うことを特徴とする、請求項１に記載の音声再生装置。

【請求項５】

入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するための音声再生方法であって、

前記音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する判別ステップと、

前記音声区間および前記非音声区間に基づき、算出用フレーム長に対して当該音声区間が含まれる比率を示す音声含有率と、前記算出用フレーム長の音声含有率の平均値および標準偏差である統計値を少なくとも算出する音声情報算出ステップと、

前記音声速度変換処理の比率が１以上の速度比を基準値として予め設定し、前記統計値および前記算出用フレーム長の音声含有率に応じて、前記基準値から前記音声区間の速度比を算出する速度比算出ステップとを備え、

前記音声情報算出ステップは、それぞれ時間長が異なる前記算出用フレーム長を複数設定してそれぞれ前記音声含有率および前記統計値を算出し、

前記速度比算出ステップは、前記算出用フレーム長における音声含有率の前記平均値に対する差分値および前記標準偏差によって前記算出用フレーム長毎に得られる値を用いて算出される係数を前記速度比の基準値に乗ずることで、前記算出用フレーム長の音声含有率が相対的に高いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より小さな値に変更し、前記算出用フレーム長の音声含有率が相対的に低いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より大きな値に変更し、音声含有率の時間的な変動に対して適応的に前記音声区間の速度比を調整する速度制御を行う音声再生方法。

【請求項６】

入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するためのコンピュータで実行される音声再生プログラムを記録した当該コンピュータで読み取り可能な記録媒体であって、

前記コンピュータに、

前記音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する

判別ステップと、
前記音声区間および前記非音声区間に基づき、算出用フレーム長に対して当該音声区間が含まれる比率を示す音声含有率と、前記算出用フレーム長の音声含有率の平均値および標準偏差である統計値を少なくとも算出する音声情報算出ステップと、
前記音声速度変換処理の比率が１以上の速度比を基準値として予め設定し、前記統計値および前記算出用フレーム長の音声含有率に応じて、前記基準値から前記音声区間の速度比を算出する速度比算出ステップとを含み、
前記音声情報算出ステップは、それぞれ時間長が異なる前記算出用フレーム長を複数設定してそれぞれ前記音声含有率および前記統計値を算出し、
前記速度比算出ステップは、前記算出用フレーム長における音声含有率の前記平均値に対する差分値および前記標準偏差によって前記算出用フレーム長毎に得られる値を用いて算出される係数を前記速度比の基準値に乗ずることで、前記算出用フレーム長の音声含有率が相対的に高いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より小さな値に変更し、前記算出用フレーム長の音声含有率が相対的に低いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より大きな値に変更し、音声含有率の時間的な変動に対して適応的に前記音声区間の速度比を調整する速度制御を行うプログラムを記録した、コンピュータに読み取り可能な記録媒体。
【請求項７】
入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するための集積回路であって、
前記音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する判別部と、
前記音声区間および前記非音声区間に基づき、算出用フレーム長に対して当該音声区間が含まれる比率を示す音声含有率と、前記算出用フレーム長の音声含有率の平均値および標準偏差である統計値を少なくとも算出する音声情報算出部と、
前記音声速度変換処理の比率が１以上の速度比を基準値として予め設定し、前記統計値および前記算出用フレーム長の音声含有率に応じて、前記基準値から前記音声区間の速度比を算出する速度比算出部とを備え、
前記音声情報算出部は、それぞれ時間長が異なる前記算出用フレーム長を複数設定してそれぞれ前記音声含有率および前記統計値を算出し、
前記速度比算出部は、前記算出用フレーム長における音声含有率の前記平均値に対する差分値および前記標準偏差によって前記算出用フレーム長毎に得られる値を用いて算出される係数を前記速度比の基準値に乗ずることで、前記算出用フレーム長の音声含有率が相対的に高いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より小さな値に変更し、前記算出用フレーム長の音声含有率が相対的に低いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より大きな値に変更し、音声含有率の時間的な変動に対して適応的に前記音声区間の速度比を調整する速度制御を行う集積回路。
【請求項８】
入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するための音声録音再生装置であって、
前記入力される音声信号を記録する情報記録部と、
前記情報記録部に記録される前の音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する判別部と、
前記音声区間および前記非音声区間に基づき、算出用フレーム長に対して当該音声区間が含まれる比率を示す音声含有率と、前記算出用フレーム長の音声含有率の平均値および標準偏差である統計値を少なくとも算出する音声情報算出部と、
前記音声速度変換処理の比率が１以上の速度比を基準値として予め設定し、前記統計値および前記算出用フレーム長の音声含有率に応じて、前記基準値から前記音声区間の速度比を算出する速度比算出部とを備え、

前記音声情報算出部は、それぞれ時間長が異なる前記算出用フレーム長を複数設定してそれぞれ前記音声含有率および前記統計値を算出し、

前記速度比算出部は、前記算出用フレーム長における音声含有率の前記平均値に対する差分値および前記標準偏差によって前記算出用フレーム長毎に得られる値を用いて算出される係数を前記速度比の基準値に乗ずることで、前記算出用フレーム長の音声含有率が相対的に高いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より小さな値に変更し、前記算出用フレーム長の音声含有率が相対的に低いときに当該算出用フレーム長における前記音声区間の速度比を当該基準値より大きな値に変更し、音声含有率の時間的な変動に対して適応的に前記音声区間の速度比を調整する速度制御を行う音声録音再生装置。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、音声再生装置、音声録音再生装置、およびそれらの方法、記録媒体、集積回路に関し、より特定的には、再生速度を変換して再生する音声再生装置、音声録音再生装置、およびそれらの方法、記録媒体、集積回路に関する。

【背景技術】

【０００２】

従来、予め記録された音声を再生する音声再生装置において、声の高さを変えることなく、より高速に再生する方法が知られている（例えば、特許文献１参照）。特許文献１に開示された音声再生装置では、音声信号全体を指定速度で再生するとき、音声区間については部分的に再生速度比を低速化している。これにより、特許文献１に開示された従来の音声再生装置は、情報の欠落が少なく、聴き取りやすい再生音声を提供することができる。

【特許文献１】特開２００１－２２２３００号公報

【０００３】

以下、図１１を参照して、上記特許文献１に開示された従来の音声再生装置９について、具体的に説明する。図１１は、従来の音声再生装置９の構成を示すブロック図である。図１１において、従来の音声再生装置９は、音響分析部９１、話速変換部９２、非音声区間長制御部９３、および合成部９４を備える。

【０００４】

音響分析部９１は、入力される音声データに対して、予め設定されているパワー閾値に基づき音声区間および非音声区間を判別する。そして、音響分析部９１は、音声区間および非音声区間の時間情報をそれぞれ求める。図１１に示す従来の音声再生装置９では、音響分析部９１において判別された音声区間および非音声区間に対して、異なる再生処理を適用する。音響分析部９１で判別された音声区間の音声データおよび上記各時間情報は、話速変換部９２に出力される。音響分析部９１で判別された非音声区間の音声データは、非音声区間長制御部９３に出力される。

【０００５】

話速変換部９２は、まず音声区間の音声データと上記各時間情報とに基づいて、一定時間長以上の非音声区間に挟まれた音声区間を特定する。そして、話速変換部９２は、当該音声区間の冒頭部分の速度比を所定速度比より遅く、末尾に向けて次第に所定速度比に戻すような速度比制御を行う。速度比が制御された音声区間の音声データは、合成部９４に出力される。また、話速変換部９２は、波形の伸長処理によって生じる音声区間の遅延時間情報を非音声区間長制御部９３に出力する。

【０００６】

一方、非音声区間長制御部９３では、話速変換部９２から出力された上記遅延時間情報に基づいて、非音声区間の音声データに対して削除および圧縮する処理を適宜行う。つまり、非音声区間長制御部９３では、目標の指定速度比に合うように、かつ、話速変換部９

２で生じた音声区間の遅延を解消するような処理が行われる。非音声区間長制御部９３において処理された非音声区間の音声データは、合成部９４に出力される。

【０００７】

合成部９４は、話速変換部９２から出力された音声区間の音声データと、非音声区間長制御部９３から出力された非音声区間の音声データとを合成する。そして、合成部９４は、速度比が変換された音声区間と非音声区間とが合成された音声データを変換音声データとして、最終的な再生音声を出力する。

【０００８】

上記従来の音声再生装置９では、例えば指定速度としてｍ倍速（ｍは１以上の正数）が与えられたとき、音声区間の冒頭部分ではｍ倍速より遅い速度比で再生する。そして、従来の音声再生装置９は、音声区間の末尾に向かって次第に再生速度比を速くする。ここで、一般的に音声区間の冒頭部分には、重要な情報が含まれている場合が多い。したがって、従来の音声再生装置９によれば、音声区間の冒頭部分にある重要な情報を欠落させることなく、聴きとりやすい再生を実現することができる。このように従来の音声再生装置９では、音声区間については聴き取りやすい処理が、非音声区間については指定速度比に適応するような処理がそれぞれ行われている。

【発明の開示】

【発明が解決しようとする課題】

【０００９】

ここで、高速再生時には、音声の発話速度が速くなり、ユーザにとって内容を理解するための負荷が大きくなる。さらに、番組全体の中で音声区間が偏って集中すると（音声が連続的に発声されると）、ユーザにとってさらに理解が困難になる。しかしながら、上記従来の音声再生装置９では、一つの音声区間の中で再生速度比を変更することのみを想定している。つまり、上記従来の音声再生装置９では、例えばテレビ番組などの全体を通して、同一の速度比制御処理が適用される。したがって、従来の音声再生装置９においては、音声区間が偏って集中する部分で相対的に音声の内容の聴き取りが困難になるという本質的課題があった。

【００１０】

それ故、本発明の目的は、テレビなどの番組全体を考慮した最適な速度比制御を行って、より聴き取りやすい再生を実現する音声再生装置、音声録音再生装置、およびそれらの方法、記録媒体、および集積回路を提供することを目的とする。

【課題を解決するための手段】

【００１１】

第１の発明は、入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するための音声再生装置であって、音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する判別部と、音声区間および非音声区間に基づき、所定時間長に対して当該音声区間が含まれる比率を示す音声含有率を少なくとも算出する音声情報算出部と、音声速度変換処理の比率が１以上の速度比を基準値として予め設定し、所定時間長の音声含有率が相対的に高いときに当該所定時間長における音声区間の速度比を当該基準値より小さな値に変更し、所定時間長の音声含有率が相対的に低いときに当該所定時間長における音声区間の速度比を当該基準値より大きな値に変更する速度比算出部とを備え、音声含有率の時間的な変動に対して適応的に音声区間の速度比を調整する速度制御を行うものである。

【００１２】

第２の発明は、上記第１の発明において、速度比算出部は、ユーザの操作に応じて短縮された再生時間を設定し、音声区間における変更した速度比に基づいて、音声信号の再生時間が設定された再生時間となるように非音声区間の速度比を算出することを特徴とする。

【００１３】

第３の発明は、上記第２の発明において、速度比算出部は、設定された再生時間内にお

いて非音声区間の速度比を一定に算出することを特徴とする。

【００１４】

第４の発明は、上記第１の発明において、速度比算出部は、所定時間長の設定値を選択することによって音声含有率の時間的な変動に対する音声区間の速度比の適応度合いを可変することを特徴とする。

【００１５】

第５の発明は、上記第１の発明において、音声再生装置は、入力される音声信号のうち、少なくとも所定時間長分の音声信号を含むように当該音声信号を順次更新しながら記録するバッファと、バッファに記録された音声信号に対して音声速度変換処理を行って出力する速度変換部とを、さらに備え、判別部は、バッファに記録された所定時間長の音声信号に対して音声区間と非音声区間とを判別し、音声情報算出部は、さらに、所定時間長の音声含有率に関する統計値を算出して、これまでに算出した統計値を単位時間毎に順次更新し、速度比算出部は、単位時間ごとに更新される統計値および当該更新時の所定時間長に設定された音声含有率に応じて音声区間の速度比を算出し、速度変換部は、バッファで順次更新される音声信号に対して、単位時間ごとに算出された音声区間の速度比を用いて順次音声速度変換処理を行うことを特徴とする。

【００１６】

第６の発明は、上記第１の発明において、音声情報算出部は、所定時間長の音声含有率に関する統計値をさらに算出し、速度比算出部は、統計値および所定時間長の音声含有率に応じて音声区間の速度比を算出することを特徴とする。

【００１７】

第７の発明は、上記第５または６の発明において、統計値は、所定時間長の音声含有率の平均値および標準偏差であることを特徴とする。

【００１８】

第８の発明は、上記第７の発明において、速度比算出部は、所定時間長における音声含有率の平均値に対する差分値および標準偏差によって得られる値を用いて算出される係数を速度比の基準値に乗じて、音声区間の速度比を算出することを特徴とする。

【００１９】

第９の発明は、上記第７の発明において、音声情報算出部は、それぞれ時間長が異なる所定時間長を複数設定してそれぞれ音声含有率および統計値を算出し、速度比算出部は、所定時間長における音声含有率の平均値に対する差分値および標準偏差によって所定時間長毎に得られる値を用いて算出される係数を速度比の基準値に乗じて算出することを特徴とする。

【００２０】

第１０の発明は、入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するための音声再生方法であって、音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する判別ステップと、音声区間および非音声区間に基づき、所定時間長に対して当該音声区間が含まれる比率を示す音声含有率を少なくとも算出する音声情報算出ステップと、音声速度変換処理の比率が１以上の速度比を基準値として予め設定し、所定時間長の音声含有率が相対的に高いときに当該所定時間長における音声区間の速度比を当該基準値より小さな値に変更し、所定時間長の音声含有率が相対的に低いときに当該所定時間長における音声区間の速度比を当該基準値より大きな値に変更する速度比算出ステップとを含む。

【００２１】

第１１の発明は、入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するためのコンピュータで実行される音声再生プログラムを記録した当該コンピュータで読み取り可能な記録媒体であって、コンピュータに、音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する判別ステップと、音声区間および非音声区間に基づき、所定時間長に対する当該音声区間が含まれる比率を示す音声含有率を少なくとも算出する音声情報算出ステップと、音声速度変換処理の

比率が１以上の速度比を基準値として予め設定し、所定時間長の音声含有率が相対的に高いときに当該所定時間長における音声区間の速度比を当該基準値より小さな値に変更し、所定時間長の音声含有率が相対的に低いときに当該所定時間長における音声区間の速度比を当該基準値より大きな値に変更する速度比算出ステップとを実行させるためのプログラムを記録した、コンピュータに読み取り可能な記録媒体である。

【００２２】

第１２の発明は、入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するための集積回路であって、音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する判別部と、音声区間および非音声区間に基づき、所定時間長に対して当該音声区間が含まれる比率を示す音声含有率を少なくとも算出する音声情報算出部と、音声速度変換処理の比率が１以上の速度比を基準値として、所定時間長の音声含有率が相対的に高いときに当該所定時間長における音声区間の速度比を当該基準値より小さな値に変更し、所定時間長の音声含有率が相対的に低いときに当該所定時間長における音声区間の速度比を当該基準値より大きな値に変更する速度比算出部とを備える。

【００２３】

第１３の発明は、入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するための音声録音再生装置であって、入力される音声信号を記録する情報記録部と、情報記録部に記録される前の音声信号に対して音声を含む音声区間と音声を含まない非音声区間とを判別する判別部と、音声区間および非音声区間に基づき、所定時間長に対して当該音声区間が含まれる比率を示す音声含有率を少なくとも算出する音声情報算出部と、音声速度変換処理の比率が１以上の速度比を基準値として予め設定し、所定時間長の音声含有率が相対的に高いときに当該所定時間長における音声区間の速度比を当該基準値より小さな値に変更し、所定時間長の音声含有率が相対的に低いときに当該所定時間長における音声区間の速度比を当該基準値より大きな値に変更する速度比算出部とを備える。

【００２４】

第１４の発明は、上記第１３の発明において、情報記録部には、音声信号が記録される際に判別部が判別した結果が記録され、音声情報算出部は、情報記録部に記録された結果に基づいて、所定時間長の音声含有率を算出することを特徴とする。

【００２５】

第１５の発明は、上記第１３の発明において、情報記録部には、音声信号が記録される際に、判別部が判別した結果および所定時間長の音声含有率が記録され、速度比算出部は、情報記録部に記録された所定時間長の音声含有率を用いて、音声区間の速度比を算出することを特徴とする。

【００２６】

第１６の発明は、入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するための音声録音再生方法であって、入力される音声信号を記録する情報記録ステップと、情報記録ステップに記録される前の音声信号に対して音声を含む音声区間と音声を含まない非音声区間と判別する判別ステップと、音声区間および非音声区間に基づき、所定時間長に対して当該音声区間が含まれる比率を示す音声含有率を少なくとも算出する音声情報算出ステップと、音声速度変換処理の比率が１以上の速度比を基準値として予め設定し、所定時間長の音声含有率が相対的に高いときに当該所定時間長における音声区間の速度比を当該基準値より小さな値に変更し、所定時間長の音声含有率が相対的に低いときに当該所定時間長における音声区間の速度比を当該基準値より大きな値に変更する速度比算出ステップとを含む。

【００２７】

第１７の発明は、入力される音声信号に音声速度変換処理を適用して通常より再生時間を短縮して当該音声信号を聴取するためのコンピュータで実行される音声録音再生プログラムを記録した記録媒体であって、コンピュータに、入力される音声信号を記録部に記録

する情報記録ステップと、記録部に記録される前の音声信号に対して音声を含む音声区間と音声を含まない非音声区間と判別する判別ステップと、音声区間および非音声区間に基づき、所定時間長に対して当該音声区間が含まれる比率を示す音声含有率を少なくとも算出する音声情報算出ステップと、音声速度変換処理の比率が１以上の速度比を基準値として、所定時間長の音声含有率が相対的に高いときに当該所定時間長における音声区間の速度比を当該基準値より小さな値に変更し、所定時間長の音声含有率が相対的に低いときに当該所定時間長における音声区間の速度比を当該基準値より大きな値に変更する速度比算出ステップとを実行させるためのプログラムを記録した、コンピュータに読み取り可能な記録媒体である。

【発明の効果】

【００２８】

第１の発明によれば、音声含有率の変動に応じた音声区間の速度比を算出することで、入力された音声信号の速度変換後の再生音質を音声含有率の変動に応じた了解性の優れたものにすることができる。

【００２９】

第２の発明によれば、設定された再生時間となるように、重要な音声情報が含まれていない非音声区間の速度比を音声区間の速度比とは別に算出することで、音声区間の速度比をユーザが聴取可能な範囲内の速度比に調整することができる

【００３０】

第３の発明によれば、重要な音声情報が含まれていない非音声区間の速度比を一定の速度比とすることで、能率のよい速度変換をした再生が可能となる。

【００３１】

第４の発明によれば、例えば所定時間長が長い場合には、調整される音声区間の速度比が音声含有率の変動に対して大局的でより正確性の高い値となる。また例えば、所定時間長が短い場合には、調整される音声区間の速度比が音声含有率の変動に対して敏感でより追従性のよい値となる。つまり、所定時間長の設定値を選択して音声含有率の変動に対する音声区間の速度比の適応度合いを可変することによって、正確性または追従性を自由に選択することができる。

【００３２】

第５の発明によれば、統計値を単位時間毎に更新することで、音声信号の入力に応じて即時に速度変換処理をして再生することができる。

【００３３】

第６の発明によれば、音声区間の速度比の算出に対して、統計値を用いることで、より実際の音声含有率の変動に即した音声区間の速度比を算出することができ、結果的に速度変換後の再生音質をより了解性のある自然なものにすることができる。

【００３４】

第７の発明によれば、音声区間の存在の偏り度合いを考慮した音声区間の速度比を算出することができる。

【００３５】

第８の発明によれば、音声区間の存在の偏り度合いに即した音声区間の速度比を算出することができる。

【００３６】

第９の発明によれば、音声含有率の敏感な変動および大局的な変動の双方に対応した最適な音声区間の速度比を算出することができる。

【００３７】

第１３の発明によれば、音声含有率の変動に応じた音声区間の速度比を算出することで、記録した音声信号の速度変換後の再生音質を音声含有率の変動に応じた了解性の優れたものにすることができる。

【００３８】

第１４の発明によれば、音声信号を記録後、速度変換した再生が行われる前までの処理

時間を判別部における処理時間分だけ短縮することができる。
【００３９】
第１５の発明によれば、音声信号を記録後、速度変換した再生が行われる前までの処理時間を判別部および音声情報算出部における処理時間分だけ短縮することができ、音声信号を記録後、即時に速度変換をした再生を行うことができる。
【発明を実施するための最良の形態】
【００４０】
（第１の実施形態）
図１を参照して、本発明における第１の実施形態に係る音声再生装置について説明する。図１は、本発明における第１の実施形態に係る音声再生装置１の構成を示すブロック図である。図１において、音声再生装置１は、音声／非音声判別部１１、音声情報算出部１２、音声情報記録部１３、速度比算出部１４、および音声速度変換部１５を有する。なお、本実施形態に係る音声再生装置１は、記録メディアなどに録音された音声信号を速度変換して再生する前に一旦、録音された音声信号全体について読み出し可能であることを想定した装置である。ここで、録音対象としては、例えばテレビやラジオ番組が挙げられる。また記録メディアは、例えば映画などが予め収録されたＤＶＤ等の記録メディアであってもよい。以下の説明では、一例として、第１の実施形態に係る音声再生装置１が、録音されたテレビ番組の音声信号に対して速度変換処理を行うとする。
【００４１】
記録メディアなどに録音された音声信号が読み出され、音声／非音声判別部１１に入力される。音声／非音声判別部１１は、入力された音声信号のパワーの包絡値や周期性などの分析を行う。そして、音声／非音声判別部１１は入力された音声信号に対して音声区間および非音声区間を時間軸上で判別する。音声信号の時間軸上で判別された音声区間および非音声区間の情報（以下、判別情報という）は、速度変換した再生を行う前に音声情報算出部１２に出力される。
【００４２】
音声情報算出部１２は、音声／非音声区間の判別情報に基づいて、音声区間および非音声区間の速度比を算出するために必要な音声情報を算出する。音声情報としては、音声含有率、音声含有率の平均値、および標準偏差などを算出する。具体的には、音声情報算出部１２は、録音された番組全体を通して音声含有率を算出した後に、音声含有率の平均値と標準偏差とを算出する。音声情報算出部１２で算出された音声含有率、音声含有率の平均値、および標準偏差は、音声情報記録部１３にそれぞれ記録される。以下、音声含有率、音声含有率の平均値、および標準偏差について説明する。
【００４３】
音声含有率は、所定数（少なくとも１つ以上）のフレームに対して音声区間が含まれる時間比率を示すものである。音声含有率はフレーム毎に算出される。ここでフレームとは、入力される音声信号を分析するための処理単位であり、当該フレームの時間長をフレーム長とする。当該フレームには、音声区間および／または非音声区間が含まれる。また、音声含有率の算出に用いられる少なくとも１つ以上のフレームを算出用フレームとし、その時間長を算出用フレーム長とする。以下の説明では、一例として、１フレームの時間長（１フレーム長）を１分とする。また、音声含有率を算出するための算出用フレーム長をｎ（ｎは正数）分とする。つまり、１フレーム長を１分としたので、算出用フレームはｎ個のフレームから構成されることとなる。また、録音された番組全体のフレーム数がＮ（Ｎは正数）個あるとする。そして、フレームナンバーをｋ（ｋ＝１～Ｎ）として、フレームナンバーがｋのときのフレームを「第ｋフレーム」とする。このとき、第ｋフレームの音声含有率Ｒｉｓ＿ｎ（ｋ）は、数式（１）で表現される。

【数１】

$$Ris\_n(k) = \frac{(\text{ｋフレームを含むそれから先のｎ分間における音声区間の時間長})}{(\text{ｋフレームを含むそれから先のｎ分間におけるフレーム長})} \quad \cdots \quad (1)$$

つまり、数式（１）によって算出される第ｋフレームの音声含有率Ｒｉｓ＿ｎ（ｋ）は、算出用フレーム長に対して音声区間が含まれる時間比率を示す。

【００４４】

ここで、図２～図４を参照して、上記音声含有率Ｒｉｓ＿ｎ（ｋ）の算出例を挙げる。図２～図４では、一例として、テレビ放送のドキュメンタリ番組（３０分間）の音声含有率を算出するとし、１分、５分、および１０分の３種類の算出用フレーム長で算出している。図２は、算出用フレーム長が１分のときの音声含有率Ｒｉｓ＿１（ｋ）の算出例を示す図である。図３は、算出用フレーム長が５分のときの音声含有率Ｒｉｓ＿５（ｋ）の算出例を示す図である。図４は、算出用フレーム長が１０分のときの音声含有率Ｒｉｓ＿１０（ｋ）の算出例を示す図である。なお、図２～図４において、横軸はフレームナンバー（ｋ）を示し、縦軸は音声含有率（％）を示す。

【００４５】

図２において、第１フレーム（ｋ＝１）の音声含有率Ｒｉｓ＿１（１）は、算出用フレーム長を１分としたので、数式（１）より第１フレームの音声含有率そのものとなる。図３においては、数式（１）より算出される第１フレームの音声含有率Ｒｉｓ＿５（１）は、図２の第１～第５フレームの音声含有率を平均したものである。図４においては、数式（１）より算出される第１フレームの音声含有率Ｒｉｓ＿１０（１）は、図２の第１～第１０フレームの音声含有率を平均したものである。

【００４６】

図２～図４に示すように、各算出用フレーム長で音声含有率の変動の様子が異なることが分かる。具体的には、算出用フレーム長が短い場合（図２）には、音声含有率のフレーム間の変動が比較的大きくなる。つまり、算出用フレーム長が短い場合には、音声含有率の実際の変動が敏感に反映されたものとなる。これに対し、図３および図４に示すように、算出用フレーム長が長くなるにつれて、音声含有率のフレーム間の変動が比較的小さくなる。これは、上述したように、算出用フレーム長が長くなるにつれて各フレームの音声含有率が平均化されるためである。つまり、算出用フレーム長が長い場合には、平均化によって小さい変動が吸収され、音声含有率の変動が大局的に反映される。また、各算出用フレーム長の分散および標準偏差も、音声含有率の変動の違いにより、異なる値となる。

【００４７】

次に音声含有率の平均値および標準偏差について説明する。音声含有率の平均値は、音声含有率Ｒｉｓ＿ｎ（ｋ）を番組全体において平均した値である。上述した図２でいえば、Ｒｉｓ＿１（１）からＲｉｓ＿１（３０）の音声含有率を平均した値である。つまり、算出用フレーム長ｎ（ｎは正数）で表現すれば、音声含有率の平均値は、Ｒｉｓ＿ｎ（１）からＲｉｓ＿ｎ（Ｎ）までの音声含有率の平均である。また、標準偏差は、音声含有率Ｒｉｓ＿ｎ（ｋ）と音声含有率の平均値とを用いて算出される値である。ここで、上記図２～図４に示した音声含有率Ｒｉｓ＿ｎ（ｋ）の値をもとに、各算出用フレーム長について、それぞれ音声含有率の平均値と標準偏差とを求めると図５に示すような値となる。図５は、各算出用フレーム長の音声含有率の平均値および標準偏差の算出結果を示す図である。図５において、算出用フレーム長が１分である音声含有率の平均値Ａ１は０．５０６と、算出用フレーム長が５分である音声含有率の平均値Ａ５は０．４９８と、算出用フレーム長が１０分である音声含有率の平均値Ａ１０は０．４８８となる。また、図５において、平均値Ａ１に対する標準偏差Ｓ１は０．１６１と、平均値Ａ５に対する標準偏差Ｓ５は０．０７３と、平均値Ａ１０に対する標準偏差Ｓ１０は０．０２８となる。

【００４８】

このように、図５に示すように、標準偏差においては、算出用フレーム長が短い場合に

は、変動が大きく（ばらつきが大きく）なるために標準偏差の値が大きくなる。算出用フレーム長が長い場合には、変動が小さく（ばらつきが小さく）なるために標準偏差の値が小さくなる。つまり、標準偏差は、算出用フレーム長の長さによって大きな影響を受ける値であり、一般的には番組全体における音声区間の存在の偏りを示す値と考えることができる。

【００４９】

次に、入力される音声信号を速度変換して再生する段階において、速度比算出部１４は、音声情報記録部１３に記録された音声情報（音声含有率、音声含有率の平均値、および標準偏差）を用いて、音声区間の存在の偏りに応じた音声区間の速度比をフレーム毎に算出する。そして、速度比算出部１４は、上記音声区間の速度比とユーザなどが入力する所望再生時間とに基づいて、非音声区間の速度比を算出する。そして、速度比算出部１４は、音声／非音声判別部１１において判別された判別情報に対して、フレーム毎の速度比を設定して音声速度変換部１５へ出力する。なお、ここでは算出された各フレームの音声区間の速度比は、当該フレーム内に存在する音声区間に一律に適用されるとする。また、非音声区間の速度比は、後述するように例えば一定の速度比でフレーム内の非音声区間に適用されるとする。

【００５０】

ここで、速度比の算出方法を説明する前に、音声の再生速度と聴き取りやすさの関係について説明する。通常の再生時間より短い時間で音声信号を聴取するために、通常の再生時間に対する再生時間長の設定値である目標再生時間比Ｒｔ（０＜Ｒｔ＜１）が与えられたとする。例えばユーザが通常の再生時間に対して半分の再生時間で聴取しようとすると、目標再生時間比ＲｔはＲｔ＝０．５となる。このような目標再生時間比Ｒｔは、数式（２）で表現される。数式（２）において、音声含有率の平均値をＡ０と、音声含有率が一定であるときの音声区間の速度比をＳＲｓ０と、および音声含有率が一定であるときの非音声区間の速度比をＳＲｎｓ０とする。

【数２】

$$Rt = \frac{A0}{SRs0} + \frac{1-A0}{SRns0} \quad \cdots (2)$$

数式（２）より、目標再生時間比Ｒｔおよび音声含有率の平均値Ａ０が与えられ、音声区間の速度比ＳＲｓ０および非音声区間の速度比ＳＲｎｓ０のうち、いずれか一方が決まれば残りの他方が算出されることが分かる。

【００５１】

数式（２）に示す音声区間の速度比ＳＲｓ０は、一般的に通常速（等倍速）である１．０に近い値ほど聴き取りやすい。音声区間の速度比ＳＲｓ０の値が大きくなるほど、単位時間当たりの情報量が増大するので、ユーザにとって聴取が難しくなる。また、音声区間の速度比ＳＲｓ０の値が２．０程度になると、ユーザが聴き取りに集中しなければ内容を理解することが困難となる。このように、音声区間の速度比ＳＲｓ０が大きい場合、長時間の聴取にかなりの困難さが生じてくる。したがって、音声区間の速度比ＳＲｓ０は、目標再生時間比Ｒｔにある程度左右されることなく、ユーザの聴取可能な範囲内で設定されるのが最適である。これに基づき、通常は音声区間の速度比ＳＲｓ０が１～１．８程度となる範囲を利用する。また、音声／非音声判別を利用しない一定速度比での再生であれば、実用上は速度比を１．３～１．５とすることが多い。

【００５２】

本実施形態においては、上記音声区間の速度比ＳＲｓ０の最適な設定範囲を考慮しつつ、上述したように標準偏差が番組全体における音声区間の存在の偏りの度合いを示すと考え、現在のフレームにおける音声含有率と音声含有率の平均値との差および標準偏差とを用いて音声区間の速度比ＳＲｓ０を可変する。すなわち、速度比ＳＲｓ０を基準値として、音声区間が集中して音声含有率が上記音声含有率の平均値より高い部分に関しては当該基準値より音声区間の速度比を小さな値に変更し、逆に音声含有率が上記音声含有率の平

均値より低い部分に関しては当該基準値より音声区間の速度比を大きな値に変更する。

【００５３】

ここで、番組全体のフレーム数をＮと、算出用フレーム長がｎ分のときの標準偏差をＳｎと、算出用フレーム長がｎ分のときの第ｋフレームにおける音声含有率をＲｉｓ＿ｎ（ｋ）と、第ｋフレームにおける音声区間の速度比をＳＲｓ（ｋ）と、算出用フレーム長がｎ分のときの音声含有率の平均値をＡｎと、算出用フレーム長ごとに異なる重み係数をＣｎと、非音声区間の速度比をＳＲｎｓと、および音声含有率が一定と仮定したときの基準値の速度比をＳＲｓ０とする。なお、非音声区間の速度比ＳＲｎｓは、ここではフレームの音声含有率に依存せず一定値とする。このとき、音声区間の存在の偏りに応じた音声区間の速度比ＳＲｓ（ｋ）は、例えば数式（３）と表現される。

【数３】

$$SRs(k) = SRs0 * (1 + Sn * Cn * (An - Ris\_n(k))) \quad \cdots (3)$$

【００５４】

さらに、音声区間の速度比ＳＲｓ（ｋ）を音声含有率の大局的な変動および短期的な変動の双方が反映した値として算出する場合には、それぞれ時間長が異なる複数種類の算出用フレーム長の音声情報を用いて算出する。つまり、複数種類の算出用フレーム長の音声情報を多重に用いて音声区間の速度比を算出する。ここで、Ｍ種類の算出用フレーム長の音声情報を用いるとすると、第ｋフレームの音声区間の速度比ＳＲｓ（ｋ）は、数式（４）となる。

【数４】

$$SRs(k) = SRs0 * (1 + \sum_{n=1}^{M} Sn * Cn * (An - Ris\_n(k))) \quad \cdots (4)$$

数式（４）において、Ｃｎは、算出用フレーム長ごとに異なる重み係数であり、各算出用フレーム長の音声含有率の偏差を音声区間の速度比ＳＲｓ０に反映させる度合いを示すものである。

【００５５】

ここで、多重の音声情報として、算出用フレーム長が１分、５分、１０分のときの各音声情報を用いたとき、音声区間の速度比ＳＲｓ（ｋ）は、数式（５）となる。

【数５】

$$SRs(k) = SRs0 * \{1 + S1 * C1 * (A1 - Ris\_1(k)) + S5 * C5 * (A5 - Ris\_5(k)) + S10 * C10 * (A10 - Ris\_10(k))\} \quad \cdots (5)$$

ここで、数式（５）により音声情報を多重に用いた速度比の算出結果の一例を図６に示す。図６は、音声情報を多重に用いた速度比の算出結果の一例を示す図である。なお、図６に示す算出例は、数式（５）においてＳＲｓ０＝１．５、Ｃ１＝１、Ｃ５＝１０、Ｃ１０＝２０として算出し、短期的変動よりも長期的な変動に重点を置いた速度比を算出することを意図した例である。また、Ａ１、Ａ５、Ａ１０、Ｓ１、Ｓ５、Ｓ１０、Ｒｉｓ＿１（ｋ）、Ｒｉｓ＿５（ｋ）、およびＲｉｓ＿１０（ｋ）は、それぞれ図２～図５に示した値である。また、図６では、数式（５）により音声情報を多重に用いた速度比の他に、数式（３）を用いて算出フレーム長（１分、５分、および１０分）に基づく音声情報から算出された各速度比を比較のために示している。

【００５６】

図６において、菱形のプロットで描かれたグラフは、音声情報を多重に用いて算出された音声区間の速度比を示す。また、丸のプロットで描かれたグラフは、算出用フレーム長が１分のときの音声情報のみを用いて算出された音声区間の速度比を示す。四角のプロットで描かれたグラフは、算出用フレーム長が５分のときの音声情報のみを用いて算出された音声区間の速度比を示す。三角のプロットで描かれたグラフは、算出用フレーム長が１０分のときの音声情報のみを用いて算出された音声区間の速度比を示す。

【００５７】
図６に示すように、音声情報を多重に用いて算出された音声区間の速度比は、それぞれ単独の算出用フレーム長の音声情報のみを用いて算出された速度比と比べて、音声含有率の短期的な変動および長期的な変動の双方が反映された値であることが分かる。つまり、多重の音声情報を用いて算出された音声区間の速度比は、番組全体を通して音声区間の存在の偏りに応じた速度比であり、聴き取りやすい速度となるよう考慮された速度比である。

【００５８】
速度比算出部１４は、上述した方法で音声区間の速度比ＳＲｓを算出後、入力される再生時間から設定される目標再生時間比Ｒｔを達成するように非音声区間の速度比ＳＲｎｓを算出する。なお、非音声区間の速度比ＳＲｎｓは、上述したように例えば可変とせず一定の速度比とする。これは、有益な情報の大部分が音声区間に含まれているため、音声区間の速度比の調整を重視したことに基づくものである。これにより、本実施形態に係る音声再生装置は、能率良い再生を実現できる。以下、非音声区間の速度比ＳＲｎｓの算出方法について説明する。

【００５９】
目標再生時間比Ｒｔは、数式（４）に基づいて算出されたフレーム毎の音声区間の速度比ＳＲｓ（ｋ）を用いて、数式（６）と表現される。なお、Ｒｉｓ（ｋ）は、音声含有率を求める算出用フレーム長の最も短いものとする。上述の例で考えると、３種類の算出用フレーム長のうち最も短いのは、１分の算出用フレーム長である。

【数６】

$$Rt = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{Ris(k)}{SRs(k)} + \frac{(1-Ris(k))}{SRns}\right\} \quad \cdots (6)$$

【００６０】
したがって、非音声区間の速度比ＳＲｎｓは、数式（６）を整理して数式（７）となる。

【数７】

$$SRns = \frac{\sum_{k=1}^{N}(1-Ris(k))}{N*Rt - \sum_{k=1}^{N}\frac{Ris(k)}{SRs(k)}} \quad \cdots (7)$$

なお、数式（７）からも分かるように、音声区間の速度比ＳＲｓ（ｋ）がフレーム毎に算出されるのに対して、非音声区間の速度比ＳＲｎｓは、フレームには依存せず（ｋには依存せず）一定速度比として算出される。ここで、非音声区間の速度比ＳＲｎｓの算出例を挙げる。例えば音声区間の速度比が１分、５分、１０分の多重な音声情報を用いて算出されるとする。また、数式（４）において、ＳＲｓ０を１．５と、重み係数をＣ１＝１、Ｃ５＝１０、Ｃ１０＝２０とする。このとき、図６に示したように、音声情報を多重に用いて算出された音声区間の速度比ＳＲｓ（ｋ）は１．２３～１．６８の範囲の値となる。ここで、目標再生時間比Ｒｔを例えば０．５とする。このとき、非音声区間の速度比ＳＲｎｓは、数式（７）より、３．１７７となる。つまり、非音声区間の速度比ＳＲｎｓは、音声区間の速度比（例えば図６に示す１．２３～１．６８）より高速の速度比に設定される。このように、速度比算出部１４は、音声情報記録部１３に記録された音声情報を用いて、音声含有率の変動に応じた音声区間の速度比をフレーム毎に算出し、非音声区間の速度比をフレームに関係なく一定の速度比で算出する。そして、算出された音声区間および非音声区間の速度比の情報は、音声速度変換部１５に出力される。

【００６１】

音声速度変換部１５は、速度比算出部１４において算出された音声区間および非音声区間の速度比の情報に基づいて、入力される記録メディアなどに録音された音声信号に対して、速度変換処理を行う。速度変換処理の方法としては、例えば入力される音声信号を時間軸上にて圧縮伸長して速度変換を行う方法などがある。しかし、この方法に限定されず、その他の公知方法を用いて速度変換処理が行われてもよい。このように、本実施形態の音声速度変換部１５において速度変換された音声信号は、音声／非音声判別部１１の判別結果と音声含有率に応じて動的に可変する速度比で変換された音声信号である。

【００６２】

次に、図７を参照して、本実施形態に係る音声再生装置１の処理の流れについて説明する。図７は、本実施形態に係る音声再生装置１の処理の流れを示すフローチャートである。図７において、まず、ユーザが例えば記録メディアに記録された番組全体の記録時間に対して目標とする再生時間を設定する（ステップＳ１）。これにより、目標再生時間比Ｒｔ（０＜Ｒｔ＜１）が設定される。次に、記録メディアなどに録音された番組全体が読み出され、音声／非音声判別部１１において、再生前に番組全体を通して音声区間および非音声区間を判別する（ステップＳ２）。そして、音声情報算出部１２において、ステップＳ２で判別された音声／非音声区間の情報に基づいて、複数種類の算出用フレーム長について音声含有率がそれぞれ算出される（ステップＳ３）。次に、音声情報算出部１２において、ステップＳ３で算出された各算出用フレーム長の音声含有率を用いて、音声含有率の平均値および標準偏差がそれぞれ算出される（ステップＳ４）。そして、ステップＳ３およびＳ４で算出された音声情報（音声含有率、音声含有率の平均値および標準偏差）が音声情報記録部１３に記録される（ステップＳ５）。ここまでが再生前に行われる処理である。番組全体を通して音声情報が算出された後、速度変換をする再生が開始される。再生される段階で、速度比算出部１４は、音声情報記録部１３に記録された音声情報に基づいて、音声区間の存在の偏りに応じた音声区間の速度比をフレーム毎に算出する（ステップＳ６）。次に、速度比算出部１４において、ステップＳ６で算出された音声区間の速度比と、ステップＳ１で設定された目標再生時間比Ｒｔとに基づいて、非音声区間の速度比が算出される（ステップＳ７）。そして、音声／非音声判別部１１において判別された音声／非音声区間の判別情報に対して、フレーム毎の速度比を設定して音声速度変換部１５へ出力する。ステップＳ７の次に、ステップＳ６およびＳ７で算出された音声区間および非音声区間の速度比の情報に基づいて、入力される記録メディアなどに録音された音声信号に対して、速度変換処理を行う（ステップＳ８）。以上で本実施形態に係る音声再生装置１の処理の流れについての説明を終了する。

【００６３】

以上のように、本実施形態に係る音声再生装置によれば、音声含有率を音声信号全体に対して算出後、統計値として音声含有率の平均値と標準偏差とを算出して番組中の音声区間の存在の偏り度合いを予め求め、これらの音声情報を用いて音声区間の速度比を算出することで、音声含有率の変動に応じて動的に可変する音声区間の速度比を算出することができる。つまり、本実施形態に係る音声再生装置は、音声が集中する部分には速度比を低減し、音声が集中していない部分には速度比を増加させる処理を行う。これにより、本実施形態に係る音声再生装置によれば、テレビ番組や映画など全体を通して音声の了解性を保つことができる。また、非音声区間の速度比は、所定の再生時間となるように音声区間の速度比に基づいて一定速度比として算出される。これにより、能率のよい再生速度での再生が可能となる。また、各算出用フレーム長の音声情報を多重して平均値などの統計値を求めることで、音声含有率の長期的な変動や短期的な変動に対して、追従性の高い、より滑らかな速度比の制御を実現することが可能となる。

【００６４】

なお、上述した速度比算出部１４では、各算出用フレーム長の音声情報を多重して音声区間の速度比ＳＲｓ（ｋ）を算出したが、これに限定されない。例えば、音声区間の速度比ＳＲｓ（ｋ）が単独の算出用フレーム長のみ用いて算出されたものでもよい。時間長が長い算出用フレーム長を用いて算出した場合には、算出された音声区間の速度比は、変化

する音声含有率に対して大局的な値であり、より正確性のある値となる。時間長が短い算出用フレーム長を用いて算出した場合には、算出された音声区間の速度比は、変動する音声含有率に対してより追従性のよい値となる。

【００６５】

また、上述した速度比算出部１４では、音声区間の速度比を算出するための音声情報として、音声含有率Ｒｉｓ＿ｎ（ｋ）、音声含有率の平均値Ａｎ、標準偏差Ｓｎを用いるとしたが、これに限定されない。例えば、上記標準偏差の代わりに、分散や偏差平均など、標準偏差と同等の統計値が用いられてもよい。つまり、音声区間の速度比を算出するための音声情報としては、音声含有率Ｒｉｓ＿ｎ（ｋ）以外に、音声含有率の平均値Ａｎおよび標準偏差と同等の統計値が含まれる。

【００６６】

また、上述した速度比算出部１４では、音声区間の速度比をフレーム毎に算出するとしたが、フレーム内の音声区間１つ１つに対して、さらに文頭、文中、文末などの区分に分け、各区分で速度比を可変してもよい。例えば、ある音声区間の文頭では、速度比算出部１４で算出された音声区間の速度比に対してやや速度比を小さくする。そして、文末になるにつれて速度比が大きくなるように設定する。これにより、重要な情報を多く含む文頭部分がユーザにとってより聴き取りやすいものとなる。このように、速度比算出部１４は、１つの音声区間中の各区分について速度比を可変するものであってもよい。

【００６７】

なお、上述した第１の実施形態で説明した音声／非音声判別部１１、音声情報算出部１２、速度比算出部１４、および音声速度変換部１５は、例えば音声信号を入力とし、音声速度変換部１５で速度変換された音声信号を出力とする一般的なコンピュータシステム等の情報処理装置で実現可能である。この場合、上述した動作をコンピュータに実行させるプログラムを所定の情報記録媒体に格納し、当該情報記録媒体に格納されたプログラムをコンピュータが読み出して実行することによって、本発明の実現が可能となる。この場合、上記情報処理装置に接続されたキーボードなどの入力部を用いて、ユーザが所望する再生時間を入力する。また、音声情報算出部１２で算出される音声情報は、例えば情報処理装置内のハードディスクなどに記録される。また、上記プログラムを格納する情報記録媒体は、例えば、ＲＯＭまたはフラッシュメモリのような不揮発性半導体メモリやＣＤ－ＲＯＭ、ＤＶＤ、あるいはそれらに類する光学式ディスク状記録媒体である。また、プログラムを他の媒体や通信回線を通じて上記情報処理装置に供給してもかまわない。また、音声情報算出部１２で算出される音声情報は情報処理装置内のハードディスクに記録されるとしたが、情報処理装置内のメモリや情報処理装置外の他の記録媒体に記録されてもよい。

【００６８】

（第２の実施形態）

図８を参照して、本発明における第２の実施形態に係る音声再生装置について説明する。図８は、本発明における第２の実施形態に係る音声再生装置２の構成を示すブロック図である。図８において、音声再生装置２は、入力バッファ２１、音声／非音声判別部１１、音声情報逐次更新部２２、速度比算出部１４、および音声速度変換部１５を有する。

【００６９】

なお、本実施形態に係る音声再生装置２は、例えばテレビ番組や映画などの音声信号全体が既に記録メディアなどに録音済みであり、録音された音声信号全体のうち一部（所定時間分）の音声信号を一時的に保存しながら逐次的に音声情報を算出して、音声信号の入力に応じて即座に速度変換した再生を行うことを想定した装置である。そのため、本実施形態に係る音声再生装置２は、上述した第１の実施形態に係る音声再生装置１に対して、入力バッファ２１を新たに有し、音声情報逐次更新部２２において音声情報を逐次更新する点で大きく異なる。以下、異なる点を中心に説明する。また、音声／非音声判別部１１、速度比算出部１４、および音声速度変換部１５は、上述した第１の実施形態と同様であるので、同一の符号を付して、詳細な説明を省略する。

【００７０】
記録メディアなどに録音された音声信号が入力バッファ２１に入力される。入力バッファ２１は、入力された音声信号を適宜バッファする。つまり、入力バッファ２１では、音声情報逐次更新部２２で音声情報を逐次更新するために必要な所定時間分の音声信号のデータが一時的に記録される。一時的に保存された所定時間分の音声信号は、音声／非音声判別部１１および音声速度変換部１５にそれぞれ出力される。音声／非音声判別部１１は、入力された所定時間分の音声信号に対して音声区間および非音声区間を判別する。音声／非音声判別部１１において判別された音声／非音声区間の情報は、音声情報逐次更新部２２および速度比算出部１４にそれぞれ出力される。
【００７１】
音声情報逐次更新部２２は、音声／非音声区間の判別情報に基づいて音声情報を逐次更新する。なお、第１の実施形態では数式（３）および数式（４）において、音声含有率Ｒｉｓ＿ｎ（ｋ）を音声信号全体について一旦算出した後に、統計値である音声含有率の平均値Ａｎおよび標準偏差Ｓｎを算出していた。これに対し、本実施形態では、音声信号の入力に応じて即座に速度変換した再生を行うために、統計値である上記音声含有率の平均値Ａｎおよび標準偏差Ｓｎの初期値を予め記録部（図示しない）などにそれぞれ記録設定して、当該統計値を記録部などに逐次記録しながら更新していく。以下、音声情報である音声含有率の平均値および標準偏差の更新方法について説明する。
【００７２】
音声含有率の平均値Ａｎは、起動に際して初期値が設定される。そして、音声含有率の平均値Ａｎは、音声信号が入力されるフレーム毎に逐次更新される。上記初期値は、例えば再生する番組のジャンルなどによって異なり、当該ジャンルに合わせて適宜設定される。例えば、頻繁にアナウンサが話す機会の多いテレビのニュース番組などの場合は、音声含有率の平均値が８５％程度となる。また、話者の話す機会が少ない様々な映像シーンを多用するドキュメンタリ番組などの場合は、音声含有率の平均値が５０％程度になる。
【００７３】
ここで、入力バッファ２１に記録される音声信号の所定時間分を例えば上述した算出用フレーム長（ｎ分）とする。そして、入力バッファ２１は、算出用フレーム長（ｎ分）分の音声信号を確保しながら、例えば１フレーム分の音声信号を順次記録更新していくとする。また、音声情報逐次更新部２２は、例えば音声／非音声判別部１１で１フレーム分の音声／非音声区間が判別される毎に、音声含有率の平均値Ａｎの逐次更新を行うとする。この場合、音声含有率の平均値Ａｎはフレーム毎に更新され、ｋフレーム目の逐次更新される音声含有率の平均値の更新値（以下、音声含有率の更新平均値とする）をＡｎ（ｋ）とする。このとき、音声含有率の更新平均値Ａｎ（ｋ）は、数式（８）で表現される。
【数８】

$$An(k) = \alpha 1 * An(k-1) + \beta 1 * Ris\_n(k) \quad \cdots (8)$$

なお、数式（８）において、α１およびβ１は音声含有率の更新平均値Ａｎ（ｋ）の更新速度を規定するパラメータである。すなわち、α１の値が大きいほどｋフレームの１つ前のフレームの更新平均値Ａｎ（ｋ－１）の占める割合が高くなり、更新平均値Ａｎ（ｋ）の更新速度が緩やかになる。また、β１の値が大きいほどｋフレームの音声含有率Ｒｉｓ＿ｎ（ｋ）の占める割合が高くなり、更新平均値Ａｎ（ｋ）の更新速度が速くなる。数値例としては、例えばα１＝０．９８、β１＝０．０２としてもよい。
【００７４】
また、標準偏差Ｓｎも上記音声含有率の平均値と同様に、起動に際して初期値が設定される。そして、標準偏差Ｓｎは、フレーム毎に逐次更新される。上記初期値は、音声含有率の平均値Ａｎと同様に、例えば再生する番組のジャンルなどによって異なり、当該ジャンルに合わせて適宜設定される。具体的には標準偏差Ｓｎは、上記初期値と、更新平均値Ａｎ（ｋ）と、ｋフレームの音声含有率Ｒｉｓ＿ｎ（ｋ）とを用いて更新される。ここで、ｋフレーム目の標準偏差の更新値をＳｎ（ｋ）とすると、標準偏差の更新値Ｓｎ（ｋ）

は、数式（９）で表現される。
【数９】

$$Sn(k) = \sqrt{\alpha 2 * \left(Sn\,(k-1)\right)^2 + \beta 2 * \left(Ris\_n\,(k) - An\,(k)\right)^2} \quad \cdots (9)$$

なお、数式（９）において、α２およびβ２は標準偏差の更新値Ｓｎ（ｋ）の更新速度を規定するパラメータである。数値例としては、例えばα２＝０．９８、β２＝０．０２としてもよい。
【００７５】
次に、速度比算出部１４は、音声含有率Ｒｉｓ＿ｎ（ｋ）と、フレーム毎に更新された音声含有率の更新平均値Ａｎ（ｋ）および標準偏差の更新値Ｓｎ（ｋ）とに基づいて、上述した第１の実施形態と同様に、数式（３）～数式（５）に基づいて音声区間の速度比ＳＲｓ（ｋ）を算出する。また、速度比算出部１４は、算出した音声区間の速度比ＳＲｓ（ｋ）と目標再生時間比Ｒｔとに基づいて非音声区間の速度比ＳＲｎｓを算出する。そして、速度比算出部１４は、音声／非音声判別部１１から入力される音声／非音声区間の判別情報に対して、フレーム毎の速度比を設定して音声速度変換部１５へ出力する。音声速度変換部１５は、速度比算出部１４において算出された音声区間および非音声区間の速度比の情報に基づいて、入力バッファ２１から入力される音声信号に対してフレーム毎に逐次速度変換処理を行う。
【００７６】
以上のように、本実施形態に係る音声再生装置２は、統計値である音声含有率の平均値および標準偏差を逐次更新する。これにより、本実施形態に係る音声再生装置２は、音声情報を番組全体に対して事前に算出することなく、音声信号の入力に応じて即時に速度変換処理を行うことができる。
【００７７】
なお、上述した第２の実施形態で説明した音声再生装置２は、音声／非音声判別部１１、音声情報逐次更新部２２、速度比算出部１４、および音声速度変換部１５は、例えば音声信号を入力とし、音声速度変換部１５で速度変換された音声信号を出力とする一般的なコンピュータシステム等の情報処理装置で実現可能である。この場合、上述した動作をコンピュータに実行させるプログラムを所定の情報記録媒体に格納し、当該情報記録媒体に格納されたプログラムをコンピュータが読み出して実行することによって、本発明の実現が可能となる。また、上記情報処理装置に接続されるキーボードなどの入力部において、ユーザが所望する再生時間や上述した初期値を入力する。また、入力バッファ２１は、例えば情報処理装置内のハードディスク内で構成される。また、上記プログラムを格納する情報記録媒体は、例えば、ＲＯＭまたはフラッシュメモリのような不揮発性半導体メモリやＣＤ－ＲＯＭ、ＤＶＤ、あるいはそれらに類する光学式ディスク状記録媒体である。また、プログラムを他の媒体や通信回線を通じて上記情報処理装置に供給してもかまわない。また、入力バッファ２１を例えば情報処理装置内のハードディスク内で構成されるとしたが、情報処理装置内のメモリや情報処理装置外の他の記録媒体で構成されてもよい。
【００７８】
（第３の実施形態）
図９を参照して、本発明における第３の実施形態に係る音声録音再生装置について説明する。図９は、本発明における第３の実施形態に係る音声録音再生装置３の構成を示すブロック図である。図９において、音声録音再生装置３は、音声／非音声判別部１１、情報記録部３１、音声情報算出部１２、音声情報記録部１３、速度比算出部１４、および音声速度変換部１５を有する。
【００７９】
なお、本実施形態に係る音声録音再生装置３は、情報記録部３１に音声を記録して再生する音声録音再生装置であって、入力される音声信号を情報記録部３１に記録すると同時に、音声／非音声判別部１１で判別された音声区間や非音声区間の情報も情報記録部３１

に記録することを特徴とする装置である。以下、この特徴を中心に説明する。また、音声／非音声判別部１１、音声情報算出部１２、音声情報記録部１３、速度比算出部１４、および音声速度変換部１５は、上述した第１の実施形態と同様であるので、同一の符号を付して、詳細な説明を省略する。

【００８０】

録音対象となる音声信号が音声／非音声判別部１１および情報記録部３１にそれぞれ入力される。音声／非音声判別部１１は、入力された音声信号に対して音声区間および非音声区間を判別する。音声／非音声判別部１１において判別された音声／非音声区間の判別情報は、情報記録部３１に出力される。情報記録部３１において、入力された録音対象である音声信号と音声／非音声区間の判別情報とがそれぞれ記録される。

【００８１】

音声情報算出部１２は、情報記録部３１に記録された音声信号全体についての音声／非音声区間の情報を読み出して、音声情報を算出する。具体的には、音声情報算出部１２は、記録された音声信号全体を通して音声含有率を算出した後に、音声含有率の平均値および標準偏差を算出する。そして、音声情報算出部１２で算出された音声含有率、音声含有率の平均値、および標準偏差は、音声情報記録部１３にそれぞれ記録される。

【００８２】

そして、再生される段階において、速度比算出部１４は、音声情報記録部１３に記録された音声情報を用いて、音声含有率の変動に応じた音声区間の速度比をフレーム毎に算出する。また、速度比算出部１４は、音声区間の速度比と目標再生時間比Ｒｔとに基づいて非音声区間の速度比を算出する。そして、記録された音声／非音声区間の判別情報に対して、フレーム毎の速度比を設定して音声速度変換部１５へ出力する。音声速度変換部１５は、速度比算出部１４において算出された音声区間および非音声区間の速度比の情報に基づいて、情報記録部３１に記録された音声信号に対して速度変換処理を行う。

【００８３】

以上のように、本実施形態に係る音声録音再生装置３は、入力される音声信号を情報記録部３１に記録するとともに、音声／非音声判別部１１で判別された音声区間や非音声区間の情報も情報記録部３１に記録している。これにより、本実施形態に係る音声録音再生装置３によれば、音声信号全体を記録した段階で音声信号全体についての音声区間や非音声区間の判別が終了しているため、再生前に行われる音声情報の算出時間を短縮することができる。

【００８４】

なお、上述した情報記録部３１において、音声／非音声判別部１１で判別された音声区間や非音声区間の判別情報に加え、さらに音声情報算出部１２で算出された音声情報が記録されてもよい。この場合、図１０に示すように、音声情報記録部１３は省略される。図１０は、情報記録部３１に音声区間や非音声区間の情報と音声情報とを記録する音声録音再生装置４の構成を示すブロック図である。図１０において、音声録音再生装置４は、音声／非音声判別部１１、情報記録部３１、音声情報算出部１２、速度比算出部１４、および音声速度変換部１５を有する。

【００８５】

図１０において、情報記録部３１では、入力された録音対象である音声信号と、音声／非音声判別部１１において判別された音声／非音声区間の情報と、音声情報算出部１２で算出された音声情報とがそれぞれ記録される。つまり、音声録音再生装置４は、記録とともに音声／非音声区間の判別情報および音声情報が情報記録部３１に記録される。これにより、音声録音再生装置４によれば、記録後において再生時間が入力されれば、即時に速度比を算出することができる。その結果、音声録音再生装置４は、速度変換した再生音声を短時間で出力することができる。

【００８６】

なお、上述した第３の実施形態で説明した音声／非音声判別部１１、音声情報算出部１２、音声情報記録部１３、速度比算出部１４、および音声速度変換部１５は、例えば音声

信号を入力とし、音声速度変換部１５で速度変換された音声信号を出力とする一般的なコンピュータシステム等の情報処理装置で実現可能である。この場合、上述した動作をコンピュータに実行させるプログラムを所定の情報記録媒体に格納し、当該情報記録媒体に格納されたプログラムをコンピュータが読み出して実行することによって、本発明の実現が可能となる。また、上記情報処理装置に接続されるキーボードなどの入力部において、ユーザが所望する再生時間が入力される。また、情報記録部３１および音声情報記録部１３は、例えば情報処理装置内のハードディスク内で構成される。また、上記プログラムを格納する情報記録媒体は、例えば、ＲＯＭまたはフラッシュメモリのような不揮発性半導体メモリやＣＤ－ＲＯＭ、ＤＶＤ、あるいはそれらに類する光学式ディスク状記録媒体である。また、プログラムを他の媒体や通信回線を通じて上記情報処理装置に供給してもかまわない。また、情報記録部３１および音声情報記録部１３を例えば情報処理装置内のハードディスク内で構成されるとしたが、情報処理装置内のメモリや情報処理装置外の他の記録媒体で構成されてもよい。

【００８７】

また、上述した第１～第３の実施形態で説明した音声／非音声判別部１１、音声情報算出部１２、音声情報記録部１３、速度比算出部１４、音声情報逐次更新部２２および音声速度変換部１５は、例えば音声信号、再生時間情報、および上述した初期値などを入力とし、音声速度変換部１５で速度変換された音声信号を出力とする集積回路でも実現可能である。この場合、第１の実施形態における音声情報記録部１３、第２の実施形態における入力バッファ２１、第３の実施形態における音声情報記録部１３および情報記録部３１は、例えば集積回路内のメモリで構成される。そして、上述した機能を果たす電気回路を１つの小型パッケージに集積して、音声信号の処理等を行う音声信号処理回路ＤＳＰ（Ｄｉｇｉｔａｌ　Ｓｉｇｎａｌ　Ｐｒｏｃｅｓｓｏｒ）等を構成することによって、本発明の実現が可能となる。なお、第１の実施形態における音声情報記録部１３、第２の実施形態における入力バッファ２１、第３の実施形態における音声情報記録部１３および情報記録部３１は、上記集積回路とは別の他の記録媒体で構成されてもよい。

【産業上の利用可能性】

【００８８】

本発明に係る音声再生装置、音声録音再生装置、およびそれらの方法、記録媒体、および集積回路は、音声含有率の変動に応じた最適な速度比制御を行って、より聴き取りやすい再生を実現するＤＶＤプレーヤ、ＨＤＤプレーヤ、ＣＤプレーヤ等にも有用である。

【図面の簡単な説明】

【００８９】

【図１】本発明における第１の実施形態に係る音声再生装置１の構成を示すブロック図
【図２】算出用フレーム長が１分のときの音声含有率Ｒｉｓ＿１（ｋ）の算出例を示す図
【図３】算出用フレーム長が５分のときの音声含有率Ｒｉｓ＿５（ｋ）の算出例を示す図
【図４】算出用フレーム長が１０分のときの音声含有率Ｒｉｓ＿１０（ｋ）の算出例を示す図
【図５】各算出用フレーム長の音声含有率の平均値および標準偏差の算出結果を示す図
【図６】多重の音声情報を用いた速度比の算出結果の一例を示す図
【図７】本実施形態に係る音声再生装置１の処理の流れを示すフローチャート
【図８】本発明における第２の実施形態に係る音声再生装置２の構成を示すブロック図
【図９】本発明における第３の実施形態に係る音声録音再生装置３の構成を示すブロック図
【図１０】情報記録部３１に音声区間や非音声区間の情報と音声情報とを記録する音声録音再生装置４の構成を示すブロック図
【図１１】従来の音声再生装置９の構成を示すブロック図

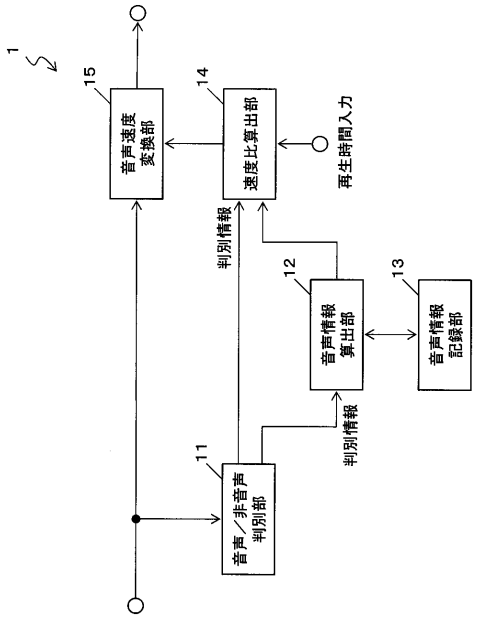
【符号の説明】

【００９０】

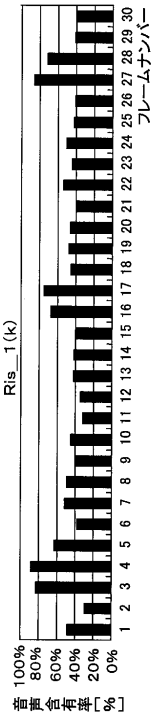
１、２　音声再生装置

３、４　音声録音再生装置
１１　音声／非音声判別部
１２　音声情報算出部
１３　音声情報記録部
１４　速度比算出部
１５　音声速度変換部
２１　入力バッファ
２２　音声情報逐次更新部
３１　情報記録部

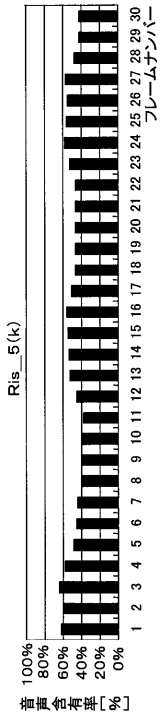
【図１】



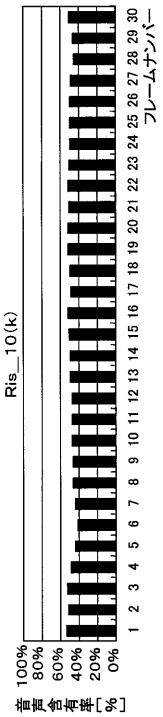
【図２】



【図３】



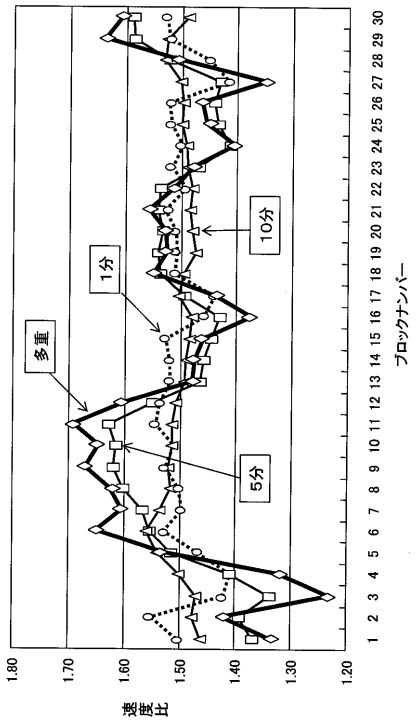
【図４】



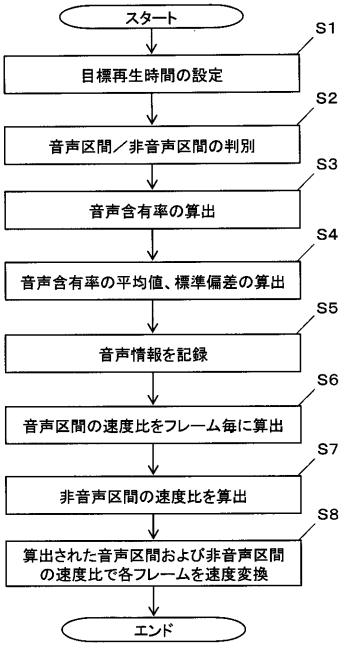
【図５】

算出用フレーム長[分]	平均値	標準偏差
1	A1； 0.506	S1； 0.161
5	A5； 0.498	S5； 0.073
10	A10； 0.488	S10； 0.028

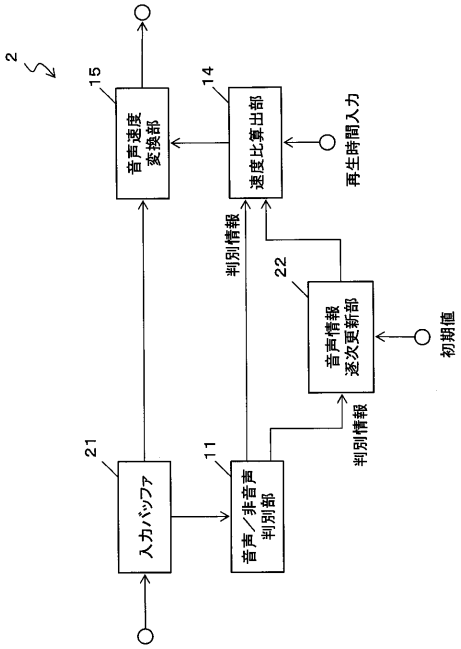
【図６】



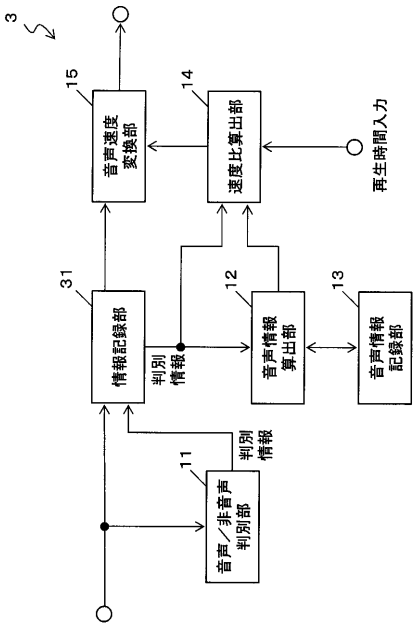
【図７】


スタート
S1
目標再生時間の設定
S2
音声区間／非音声区間の判別
S3
音声含有率の算出
S4
音声含有率の平均値、標準偏差の算出
S5
音声情報を記録
S6
音声区間の速度比をフレーム毎に算出
S7
非音声区間の速度比を算出
S8
算出された音声区間および非音声区間の速度比で各フレームを速度変換
エンド

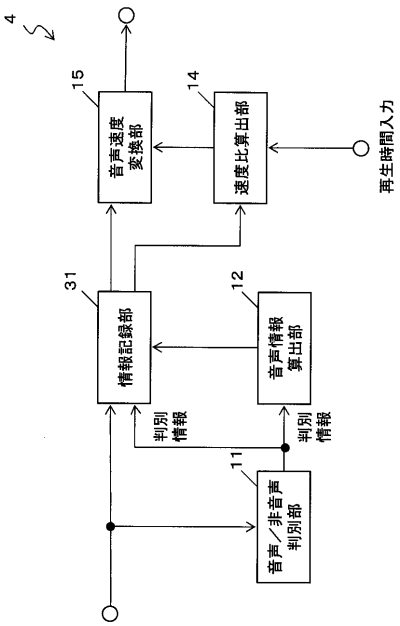
【図８】


2
15
音声速度変換部
14
速度比算出部
再生時間入力
判別情報
22
音声情報逐次更新部
初期値
判別情報
21
入力バッファ
11
音声／非音声判別部

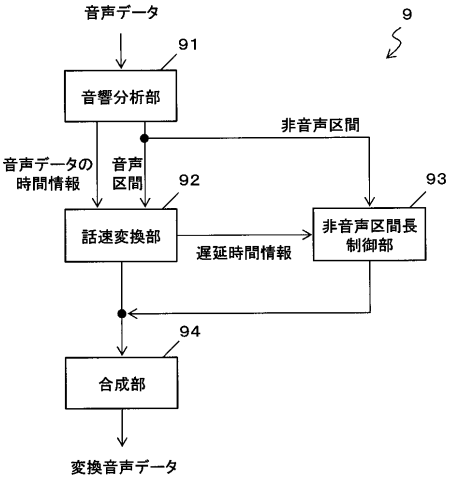
【図９】


3
15
音声速度変換部
14
速度比算出部
再生時間入力
31
情報記録部
判別情報
12
音声情報算出部
13
音声情報記録部
判別情報
11
音声／非音声判別部

【図１０】


4
15
音声速度変換部
14
速度比算出部
再生時間入力
31
情報記録部
12
音声情報算出部
判別情報
判別情報
11
音声／非音声判別部

【図１１】


音声データ
9
91
音響分析部
非音声区間
音声データの
時間情報
音声
区間
92
93
話速変換部
非音声区間長
制御部
遅延時間情報
94
合成部
変換音声データ

フロントページの続き

審査官　安田　勇太

(56)参考文献　特開平０４－３６７８９８（ＪＰ，Ａ）
特開２００１－２２２３００（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　２１／０４
Ｇ１０Ｌ　１９／００
Ｇ１０Ｌ　１１／００