(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4965371号
(P4965371)
(45)発行日　平成24年7月4日(2012.7.4)　(24)登録日　平成24年4月6日(2012.4.6)

(51)Int.Cl.　　F I
G10L 21/04 (2006.01)　　G10L 21/04　110Z
G10L 19/00 (2006.01)　　G10L 19/00　312E

請求項の数 3　(全 45 頁)

(21)出願番号　特願2007-195708(P2007-195708)
(22)出願日　平成19年7月27日(2007.7.27)
(65)公開番号　特開2008-58956(P2008-58956A)
(43)公開日　平成20年3月13日(2008.3.13)
審査請求日　平成22年7月27日(2010.7.27)
(31)優先権主張番号　特願2006-208689(P2006-208689)
(32)優先日　平成18年7月31日(2006.7.31)
(33)優先権主張国　日本国(JP)

(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　110001276
特許業務法人　小笠原特許事務所
(72)発明者　前田　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　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　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

(57)【特許請求の範囲】
【請求項１】
入力されるオーディオ信号を所定の再生時間で再生するために前記オーディオ信号に速度変換処理を適用して圧縮伸長するための目標圧伸比が設定される音声再生装置であって、
前記オーディオ信号に対して、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手段と、
前記判別手段において判別された前記音声区間に基づいて、前記オーディオ信号中に含まれる音声区間の時間比率を示す音声含有率を算出する音声含有率算出手段と、
（１）前記音声含有率と、
（２）前記目標圧伸比と、
（３）前記音声区間の平均速度比と前記非音声区間の平均速度比とが満たすべき算出条件を示す複数の音声速度比算出パターンを有する音声速度比算出分布と
を用いて、前記音声区間の平均速度比と、前記非音声区間の平均速度比とを算出し、算出したそれぞれの平均速度比を速度比条件として設定する速度比条件設定手段と、
前記速度比条件設定手段により設定された前記速度比条件に基づいて、前記音声区間及び前記非音声区間のそれぞれの平均圧伸比に対する圧伸比変化量の和が、前記音声区間又は前記非音声区間内においてゼロとなるように、前記音声区間の平均速度比及び前記非音声区間の平均速度比を決定する速度比決定手段と、
前記オーディオ信号に含まれる音声区間及び非音声区間の再生速度を前記速度比決定手

段で決定された前記音声区間の速度比および前記非音声区間の速度比に基づいてそれぞれ変換する速度変換手段とを備え、
前記複数の音声速度比算出パターンは、互いに異なる前記算出条件を有し、
前記速度比条件設定手段は、前記音声含有率と前記目標圧伸比に応じて、前記複数の音声速度比算出パターンの中から一つの音声速度比算出パターンが定まる前記音声速度比算出分布に基づいて、前記速度比条件を設定することを特徴とし、
前記判別手段は、予め定められた特定音を含む特定イベント区間と、前記特定音を含まない非特定イベント区間とを判別し、前記非特定イベント区間に対して前記音声区間と前記非音声区間とを判別し、
前記判別手段により判別された前記特定イベント区間に基づいて、前記オーディオ信号中に含まれる特定イベント区間の時間比率を示す特定イベント含有率を算出する特定イベント含有率算出手段をさらに備え、
前記速度比条件設定手段は、
（４）前記特定イベント含有率と、
（５）前記目標圧伸比と、
（６）前記特定イベント区間の平均速度比と前記非特定イベント区間の平均速度比が満たすべき算出条件を示す複数の特定イベント速度比算出パターンを有する特定イベント速度比算出分布とを用いて、
前記特定イベント区間の平均速度比と前記非特定イベント区間の平均速度比を算出し、算出したそれぞれの平均速度比を前記速度比条件としてさらに設定し、
前記複数の特定イベント速度比算出パターンは、互いに異なる前記算出条件を有し、
前記速度比条件設定手段は、前記特定イベント含有率と前記目標圧伸比に応じて、前記複数の特定イベント速度比算出パターンの中から一つの特定イベント速度比算出パターンが定まる前記特定イベント速度比算出分布に基づいて、前記速度比条件を設定する
ことを特徴とする、音声再生装置。
【請求項２】
入力されるオーディオ信号を所定の再生時間で再生するために前記オーディオ信号に速度変換処理を適用して圧縮伸長するための目標圧伸比が設定される音声再生装置であって、
前記オーディオ信号に対して、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手段と、
前記判別手段において判別された前記音声区間に基づいて、前記オーディオ信号中に含まれる音声区間の時間比率を示す音声含有率を算出する音声含有率算出手段と、
（１）前記音声含有率と、
（２）前記目標圧伸比と、
（３）前記音声区間の平均速度比と前記非音声区間の平均速度比とが満たすべき算出条件を示す複数の音声速度比算出パターンを有する音声速度比算出分布と
を用いて、前記音声区間の平均速度比と、前記非音声区間の平均速度比とを算出し、算出したそれぞれの平均速度比を速度比条件として設定する速度比条件設定手段と、
前記速度比条件設定手段により設定された前記速度比条件に基づいて、前記音声区間及び前記非音声区間のそれぞれの平均圧伸比に対する圧伸比変化量の和が、前記音声区間又は前記非音声区間内においてゼロとなるように、前記音声区間の平均速度比及び前記非音声区間の平均速度比を決定する速度比決定手段と、
前記オーディオ信号に含まれる音声区間及び非音声区間の再生速度を前記速度比決定手段で決定された前記音声区間の速度比および前記非音声区間の速度比に基づいてそれぞれ変換する速度変換手段とを備え、
前記複数の音声速度比算出パターンは、互いに異なる前記算出条件を有し、
前記速度比条件設定手段は、前記音声含有率と前記目標圧伸比に応じて、前記複数の音声速度比算出パターンの中から一つの音声速度比算出パターンが定まる前記音声速度比算出分布に基づいて、前記速度比条件を設定することを特徴とし、

前記オーディオ信号の内容を示すコンテンツ情報を取得するコンテンツ取得手段を、さらに備え、
前記速度比条件設定手段は、
前記コンテンツ情報が示す内容に応じた複数の前記音声速度比算出分布を有し、
前記コンテンツ取得手段により取得された前記コンテンツ情報に基づいて前記複数の音声速度比算出分布の中から一つを選択し、
選択した前記音声速度比算出分布に基づいて、前記速度比条件を設定する
ことを特徴とし、
前記判別手段は、予め定められた特定音を含む特定イベント区間と、前記特定音を含まない非特定イベント区間とを判別し、前記非特定イベント区間に対して前記音声区間と前記非音声区間とを判別し、
前記判別手段により判別された前記特定イベント区間に基づいて、前記オーディオ信号中に含まれる特定イベント区間の時間比率を示す特定イベント含有率を算出する特定イベント含有率算出手段をさらに備え、
前記速度比条件設定手段は、
（４）前記特定イベント含有率と、
（５）前記目標圧伸比と、
（６）前記特定イベント区間の平均速度比と前記非特定イベント区間の平均速度比が満たすべき算出条件を示す複数の特定イベント速度比算出パターンを有する特定イベント速度比算出分布とを用いて、
前記特定イベント区間の平均速度比と前記非特定イベント区間の平均速度比を算出し、算出したそれぞれの平均速度比を前記速度比条件としてさらに設定し、
前記複数の特定イベント速度比算出パターンは、互いに異なる前記算出条件を有し、
前記速度比条件設定手段は、前記特定イベント含有率と前記目標圧伸比に応じて、前記複数の特定イベント速度比算出パターンの中から一つの特定イベント速度比算出パターンが定まる前記特定イベント速度比算出分布に基づいて、前記速度比条件を設定する
ことを特徴とする、音声再生装置。
【請求項３】
前記判別手段は、複数種類の前記特定音をそれぞれ含む複数の前記特定イベント区間と、前記複数種類の特定音を全て含まない非特定イベント区間とを判別し、
前記速度比条件設定手段は、
前記複数の特定イベント区間に対応する複数の前記特定イベント含有率にそれぞれ対応する複数の前記特定イベント速度比算出分布を有し、
判別された特定イベント区間に対応する前記複数の特定イベント速度比算出分布に基づいて、前記速度比条件を設定する
ことを特徴とする、請求項１または２に記載の音声再生装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声再生装置に関し、より特定的には、オーディオ信号の再生速度を変えて再生する音声再生装置に関する。
【背景技術】
【０００２】
従来、オーディオ信号の再生速度を変えて再生する音声再生装置として、音声を含む音声区間の再生速度と、音声を含まない非音声区間の再生速度とを別々に変える音声再生装置が提案されている（例えば、特許文献１参照）。以下、図３３を参照して従来の音声再生装置について説明する。図３３は、従来の音声再生装置の構成を示したブロック図である。
【０００３】
図３３に示す従来の音声再生装置において、ユーザがオーディオ信号全体の再生時間に

対して目標時間を設定する。この目標時間は、オーディオ信号全体を等倍の再生速度で再生したときの再生時間よりも短い時間とする。音響分析部９１は、入力されるオーディオ信号を音声区間及び非音声区間に分離する。速度変換部９２は、一定時間長以上の非音声区間に挟まれた音声区間のオーディオ信号に対して、その冒頭部分が所定の再生速度よりも遅くなり、かつ末尾に向けて次第に所定の再生速度に戻るように速度変換を行っている。ここで、話速変換部９２における上記速度変換処理によって、音声区間の再生時間が長くなり、結果的にオーディオ信号全体の再生時間が目標時間に対して遅延してしまうという問題があった。そこで、非音声区間長制御部９３は、速度変換部９２から出力される遅延時間情報を参照して、非音声区間に対して当該遅延時間を短くするための処理を行う。具体的には、非音声区間長制御部９３は、非音声区間を削除したり、圧縮したりする処理を行って、遅延時間を短くしている。速度変換部９２で速度変換された音声区間のオーディオ信号と、非音声区間長制御部９３で処理された非音声区間のオーディオ信号は、合成部９４で合成され、合成部９４から出力される。

【特許文献１】特開２００１－２２２３００号公報（第１－２頁、図１）

【発明の開示】

【発明が解決しようとする課題】

【０００４】

ここで、入力されるオーディオ信号に含まれる音声区間の比率は、入力されるオーディオ信号に応じて異なっている。しかしながら、従来の音声再生装置では、音声区間が含まれる比率に関わらず、音声区間に対しては上記速度変換処理を一律に行い、非音声区間に対しては目標時間を達成するための削除や圧縮を行っている。したがって、例えば入力されるオーディオ信号が音声区間を多く含む信号である場合、音声区間に対しては上記速度変換処理が一律に行われるので、話速変換部９２において生じる遅延時間が長くなってしまう。そして遅延時間が長くなれば、非音声区間に対する区間の削除量や圧縮量も大きくなってしまい、情報の欠落が大きくなったり、再生が聞き取り難くなったりする。このように従来の音声再生装置では、目標時間を達成しつつ、入力されるオーディオ信号に応じた適切な速度変換を行うことができなかった。

【０００５】

それ故、本発明の目的は、目標時間を達成しつつ、入力されるオーディオ信号に応じた適切な速度変換を行うことが可能な音声再生装置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】

【０００６】

第１の発明は、入力されるコンテンツのオーディオ信号の再生速度を変えて所定の再生時間で再生する音声再生装置であって、オーディオ信号に対して、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手段と、判別手段において判別された判別結果に基づいて、オーディオ信号に含まれる音声区間の比率を示す音声含有率を算出する音声含有率算出手段と、オーディオ信号の再生時間が所定の再生時間となるように、オーディオ信号に予め設定された再生速度に対する音声区間及び非音声区間の速度比を音声含有率に基づいてそれぞれ算出する速度比算出手段と、オーディオ信号を入力とし、当該オーディオ信号に含まれる音声区間及び非音声区間の再生速度を速度比に基づいてそれぞれ変換する速度変換手段とを備える。

【０００７】

第２の発明は、上記第１の発明において、速度比算出手段は、オーディオ信号に予め設定された再生速度で再生される再生時間を所定の再生時間に圧伸する比率を示す圧伸比と、音声含有率と、音声区間の平均速度比を示す音声平均速度比及び非音声区間の平均速度比を示す非音声平均速度比の算出方法との対応を示す対応情報を用いて、音声平均速度比及び非音声平均速度比をそれぞれ算出して速度比条件として設定する速度比条件設定手段と、音声区間が細分化された各区間における速度比を音声平均速度比に基づく速度比に決定するとともに、非音声区間が細分化された各区間における速度比を非音声平均速度比に基づく速度比に決定して音声区間及び非音声区間の速度比をそれぞれ算出する速度比決定

手段とを有する。

【０００８】

第３の発明は、上記第２の発明において、音声再生装置は、判別手段において判別された各音声区間の開始時刻から終了時刻までの時間を音声区間長としてそれぞれ算出する音声区間長算出手段をさらに備え、速度比条件設定手段は、音声平均速度比に応じた音声区間の終了時刻における終了速度比を速度比条件としてさらに設定し、速度比決定手段は、判別手段において判別された各音声区間に対して、音声平均速度比、音声区間長、及び終了速度比に基づいて音声区間の各区間における速度比を決定する。

【０００９】

第４の発明は、上記第２の発明において、速度比決定手段は、判別手段において判別された各音声区間に対して、音声区間の開始時刻から経過した時間を音声区間長で除算して得られる経過割合に応じて音声区間の各区間における速度比をそれぞれ決定する。

【００１０】

第５の発明は、上記第２の発明において、速度比決定手段は、音声区間の開始時刻から時間が経過するにつれて再生速度が速くなるように、音声区間の各区間における速度比を決定する。

【００１１】

第６の発明は、上記第２の発明において、速度比条件設定手段は、音声平均速度比及び非音声平均速度比の算出方法を少なくとも１種類含む対応情報であって、オーディオ信号によって構成されるコンテンツの種類に応じて異なる対応を示す対応情報を用いて、音声平均速度比及び非音声平均速度比を算出する。

【００１２】

第７の発明は、上記第２の発明において、速度比条件設定手段は、音声平均速度比及び非音声平均速度比がユーザによって指定された範囲内となるように対応情報を作成し、当該対応情報を用いて音声平均速度比及び非音声平均速度比を算出する。

【００１３】

第８の発明は、上記第２の発明において、対応情報は、音声含有率の大きさに応じて当該音声含有率と音声平均速度比及び非音声平均速度比の算出方法とが異なる対応を示す情報である。

【００１４】

第９の発明は、上記第２の発明において、速度比条件設定手段は、音声平均速度比及び非音声平均速度比の算出方法を少なくとも１種類含む対応情報であって、ユーザの使用目的に応じて異なる対応を示す対応情報を用いて、音声平均速度比及び非音声平均速度比を算出する。

【００１５】

第１０の発明は、上記第２の発明において、対応情報は、圧伸比の大きさに応じて当該圧伸比と音声平均速度比及び非音声平均速度比の算出方法とが異なる対応を示す情報である。

【００１６】

第１１の発明は、上記第１の発明において、コンテンツ全体を構成するオーディオ信号と、当該コンテンツ全体を構成するオーディオ信号に対して判別手段において判別された判別結果とを予め蓄積する蓄積手段をさらに備え、音声含有率算出手段は、蓄積手段に予め蓄積された判別結果に基づいて、コンテンツ全体を構成するオーディオ信号に含まれる音声区間の比率を示す音声含有率を算出する。

【００１７】

第１２の発明は、上記第１の発明において、音声含有率算出手段は、判別手段において過去に判別された判別結果に基づいて、速度比算出手段が算出するときに用いる音声含有率を逐次算出する。

【００１８】

第１３の発明は、上記第１２の発明において、音声含有率算出手段は、速度比算出手段

が算出するときから第１の所定時間分だけ過去に判別された判別結果に基づいて、速度比算出手段が算出するときに用いる音声含有率を第１の所定時間以下の時間である第２の所定時間毎に逐次算出し、音声再生装置は、速度変換手段に入力されるデータ量及び速度変換手段から出力されるデータ量と、オーディオ信号に予め設定された再生速度で再生される再生時間を所定の再生時間に圧伸する比率を示す圧伸比とに基づいて、第２の所定時間毎の圧伸比を逐次算出する圧伸比算出手段をさらに備え、速度比算出手段は、第２の所定時間毎の圧伸比と、第２の所定時間毎の音声含有率と、音声区間の平均速度比を示す音声平均速度比及び非音声区間の平均速度比を示す非音声平均速度比の算出方法との対応を示す対応情報を用いて、音声平均速度比及び非音声平均速度比をそれぞれ算出し、算出した音声平均速度比及び非音声平均速度比を速度比条件として第２の所定時間毎に逐次設定する速度比条件設定手段と、第２の所定時間内に含まれる音声区間及び非音声区間に対して、音声区間が細分化された各区間における速度比を音声平均速度比に基づく速度比に決定するとともに、非音声区間が細分化された各区間における速度比を非音声平均速度比に基づく速度比に決定して、第２の所定時間毎に逐次設定される速度比条件に基づいて音声区間及び非音声区間の速度比をそれぞれ算出する速度比決定手段とを有し、速度変換手段は、オーディオ信号に含まれる音声区間及び非音声区間の再生速度を速度比決定手段において算出された音声区間及び非音声区間の速度比に基づいてそれぞれ変換する。

【００１９】

第１４の発明は、上記第１３の発明において、音声再生装置は、速度比決定手段において決定される音声区間の各区間における速度比が示す第２の所定時間毎の変化を抑制するための統計量を算出する統計量算出手段をさらに備え、速度比条件設定手段は、統計量、第２の所定時間毎の音声含有率、及び第２の所定時間毎の圧伸比に基づいて音声平均速度と当該音声平均速度に基づく音声区間の終了速度比とを算出し、算出した音声平均速度及び終了速度比を速度比条件として第２の所定時間毎に逐次設定する。

【００２０】

第１５の発明は、上記第１４の発明において、統計量算出手段は、コンテンツの開始時刻から速度比算出手段が算出するときまでの判別結果に基づいて、コンテンツの開始時刻から速度比算出手段が算出するときまでに含まれる音声区間の比率を示す音声含有率を統計量として算出する。

【００２１】

第１６の発明は、上記第１３の発明において、音声再生装置は、判別手段によって過去に判別された判別結果に基づいて、速度比算出手段が算出するときに用いる音声区間の開始時刻から終了時刻までの時間である音声区間長を逐次算出する音声区間長算出手段をさらに備え、速度比条件設定手段は、音声平均速度比に応じた音声区間の終了時刻における終了速度比を速度比条件としてさらに第２の所定時間毎に逐次設定し、速度比決定手段は、判別手段において判別された各音声区間に対して、音声平均速度比、音声区間長、及び終了速度比に基づいて音声区間の各区間における速度比を決定する。

【００２２】

第１７の発明は、上記第１６の発明において、速度比算出手段が算出するときに用いる音声区間長は、判別手段において過去に判別された音声区間の開始及び終了時刻から算出される音声区間長のうち、所定区間長以上の音声区間長のみに基づいて算出される。

【００２３】

第１８の発明は、上記第１６の発明において、速度比算出手段が算出するときに用いる音声区間長は、判別手段において過去に判別された音声区間の開始及び終了時刻から算出される音声区間長の最大値に基づいて算出される。

【００２４】

第１９の発明は、上記第１の発明において、所定時間分のオーディオ信号と、当該所定時間分のオーディオ信号に対して判別手段において判別された判別結果とを予め蓄積する蓄積手段をさらに備え、音声含有率算出手段は、蓄積手段に予め蓄積された判別結果に基づいて、所定時間分のオーディオ信号に含まれる音声区間の比率を示す音声含有率を算出

する。

【００２５】

第２０の発明は、上記第１の発明において、判別手段は、オーディオ信号に対して、特定イベント音を含む特定イベント区間と、当該特定イベント区間以外の音声区間及び非音声区間とを判別し、音声再生装置は、判別手段において判別された判別結果に基づいて、オーディオ信号に含まれる特定イベント区間の比率を示す特定イベント含有率を算出する特定イベント含有率算出手段をさらに備え、速度比算出手段は、オーディオ信号に予め設定された再生速度に対する特定イベント区間の速度比を特定イベント含有率に基づいて算出するとともに、オーディオ信号の再生時間が所定の再生時間となるように、オーディオ信号に予め設定された再生速度に対する特定イベント区間以外の音声区間及び非音声区間の速度比を音声含有率に基づいてそれぞれ算出し、速度変換手段は、オーディオ信号に含まれる特定イベント区間と当該特定イベント区間以外の音声区間及び非音声区間との再生速度を速度比に基づいて変換する。

【００２６】

第２１の発明は、入力されるコンテンツのオーディオ信号の再生速度を変えて所定の再生時間で再生する音声再生方法であって、オーディオ信号に対して、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、判別ステップにおいて判別された判別結果に基づいて、オーディオ信号に含まれる音声区間の比率を示す音声含有率を算出する音声含有率算出ステップと、オーディオ信号の再生時間が所定の再生時間となるように、オーディオ信号に予め設定された再生速度に対する音声区間及び非音声区間の速度比を音声含有率に基づいてそれぞれ算出する速度比算出ステップと、オーディオ信号を入力とし、当該オーディオ信号に含まれる音声区間及び非音声区間の再生速度を速度比に基づいてそれぞれ変換する速度変換ステップとを含む。

【００２７】

第２２の発明は、入力されるコンテンツのオーディオ信号の再生速度を変えて所定の再生時間で再生する音声再生装置のコンピュータに実行させるためのプログラムであって、オーディオ信号に対して、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、判別ステップにおいて判別された判別結果に基づいて、オーディオ信号に含まれる音声区間の比率を示す音声含有率を算出する音声含有率算出ステップと、オーディオ信号の再生時間が所定の再生時間となるように、オーディオ信号に予め設定された再生速度に対する音声区間及び非音声区間の速度比を音声含有率に基づいてそれぞれ算出する速度比算出ステップと、オーディオ信号を入力とし、当該オーディオ信号に含まれる音声区間及び非音声区間の再生速度を速度比に基づいてそれぞれ変換する速度変換ステップとを、コンピュータに実行させるプログラムである。

【００２８】

第２３の発明は、上記第２２の発明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。

【００２９】

第２４の発明は、入力されるコンテンツのオーディオ信号の再生速度を変えて所定の再生時間で再生する集積回路であって、オーディオ信号に対して、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手段と、判別手段において判別された判別結果に基づいて、オーディオ信号に含まれる音声区間の比率を示す音声含有率を算出する音声含有率算出手段と、オーディオ信号の再生時間が所定の再生時間となるように、オーディオ信号に予め設定された再生速度に対する音声区間及び非音声区間の速度比を音声含有率に基づいてそれぞれ算出する速度比算出手段と、オーディオ信号を入力とし、当該オーディオ信号に含まれる音声区間及び非音声区間の再生速度を速度比に基づいてそれぞれ変換する速度変換手段とを備える。

【発明の効果】

【００３０】

上記第１の発明によれば、入力されるオーディオ信号の音声含有率を算出することによ

り、所定の再生時間を達成しつつ、入力されるオーディオ信号に対して当該オーディオ信号の音声含有率に基づいた最適な音声区間及び非音声区間の速度比をそれぞれ算出することができる。つまり、音声含有率を算出することで、入力されるオーディオ信号に含まれる音声区間の比率を知ることができ、音声区間及び非音声区間の両方について所定の再生時間を達成するための最適な速度比を算出することができる。これにより、どのようなオーディオ信号が入力されても、再生内容の不連続性や情報の欠落による不快感などを低減させた、聞き取り易い再生を実現することができる。

【００３１】

上記第２の発明によれば、所定の再生時間に圧伸する比率を示す圧伸比と音声含有率とに基づく音声平均速度比及び非音声平均速度比をそれぞれ算出して、音声区間及び非音声区間の各区間における速度比が決定されることで、所定の再生時間を達成しつつ、入力されるオーディオ信号の音声含有率に基づいた最適な音声区間及び非音声区間の速度比をそれぞれ算出することができる。

【００３２】

上記第３の発明によれば、音声平均速度比、音声区間長、及び終了速度比に基づいて音声区間の各区間における速度比が決定されることで、音声区間の終了時刻における速度比を終了速度比で一定にしつつ、文頭から文末まで音声区間長に適した速度比を決定することができ、音声区間末の再生の高速化による聞き取り難さや不自然さを低減することができる。

【００３３】

上記第４の発明によれば、音声区間の各区間における速度比を経過割合に応じて決定することで、音声区間長の長短に関わらず、簡単な関数を用いて音声区間の各区間における速度比を決定することができる。

【００３４】

上記第５の発明によれば、音声区間の冒頭部分の再生速度が他の部分と比べて相対的に遅くなるので、冒頭部分の聞き逃しによって再生内容の理解度が低下することを防ぐことができる。

【００３５】

上記第６の発明によれば、コンテンツの種類に応じて異なる音声平均速度比及び非音声平均速度比を算出することができ、音声区間及び非音声区間の各区間における速度比をコンテンツに応じたより精度の高いものにすることができる。

【００３６】

上記第７の発明によれば、音声区間及び非音声区間の各区間における速度比がユーザによって指定された範囲内の音声平均速度比及び非音声平均速度比に基づく速度比となり、ユーザの聞き取り能力や好みに応じた速度変換処理を行うことができる。

【００３７】

上記第８の発明によれば、所定の再生時間を示す圧伸比を達成しつつ、入力されるオーディオ信号の音声含有率に基づいた最適な音声区間及び非音声区間の速度比をそれぞれ算出することができる。

【００３８】

上記第９の発明によれば、ユーザの使用目的に応じて異なる音声平均速度比及び非音声平均速度比の算出方法を変更することが可能になり、早聞き再生や遅聞き再生だけではなく、挿入や一時停止等を含んだオーディオ信号の出力時間に関する様々な制御について取り扱うことができる。これにより、コンテンツを視聴用、概要把握用、語学学習用、書き起こし用など用途に分けて個別に作成する必要がなく、同一のコンテンツを様々な目的で利用可能となる。

【００３９】

上記第１０の発明によれば、圧伸比の大きさに応じて音声平均速度比及び非音声平均速度比の算出方法との対応が異なることで、例えば圧伸比が低い場合は概要把握用途に、圧伸比が高い場合は学習用途等に、ユーザの目的に応じた速度比を決定することができる。

これにより、ユーザは目的に応じて機器の使い分けを意識せずに使用でき、またユーザの視聴要望に即した速度変換処理を行うことができる。

【００４０】

上記第１１の発明によれば、コンテンツ全体についての音声含有率を算出することにより、精度の高い音声区間及び非音声区間の速度比を算出することができる。

【００４１】

上記第１２の発明によれば、蓄積手段を設けることなく処理が可能なため、オーディオ信号が蓄積手段に蓄積されるのを待つ必要がなく、リアルタイムで速度変換処理を行うことができる。

【００４２】

上記第１３の発明によれば、第１の所定時間分の音声含有率が第２の所定時間に反映されることとなり、音声含有率の変動をすぐに反映した音声区間及び非音声区間の速度比の算出が可能になる。また、音声含有率を第１の所定時間分から算出することで再生時間に誤差が生じた場合であっても、第２の所定時間毎の圧伸比を用いて速度変換処理を行うので、当該誤差をこれから先の速度変換処理において解消させることができる。

【００４３】

上記第１４の発明によれば、音声区間の各区間における速度比が示す前記第２の所定時間毎の変化を抑制するための統計量を用いることで、第１の所定時間分の音声含有率が局所的に高くなった場合でも、音声区間の各区間における速度比が上がりすぎることを防ぐことができる。その結果、音声含有率が異なる様々なコンテンツに対応した速度変換処理が可能となる。

【００４４】

上記第１５の発明によれば、第１の所定時間分の音声含有率と時間変化の傾向が異なる音声含有率を統計量として利用することで、第１の所定時間分の音声含有率が局所的に変動した場合であっても、その変動は抑制され、聞き取り易い速度変換処理が可能となる。

【００４５】

上記第１６の発明によれば、音声区間長が逐次算出されることで、音声区間の終了時刻が分からなくても、音声区間に対して適切な速度変換処理を行うことができ、より精度の高いリアルタイム処理を実現することができる。

【００４６】

上記第１７の発明によれば、所定区間長以上の音声区間長のみに基づいて算出されることで、「はい」や「うん」など相槌や、「えー」などのフィラーなどを除いた平均的な音声区間長を算出することができる。

【００４７】

上記第１８の発明によれば、速度比算出手段が算出するときに用いる音声区間長が過去の音声区間長の最大値に基づき算出されることで、音声区間の終了時刻では終了速度比で変換される割合が低下し、音声区間の平均速度比が更に下がる効果があり、聞き易い速度変換処理を実現することができる。

【００４８】

上記第１９の発明によれば、所定時間分のオーディオ信号単位で速度変換を行うことができる。これにより、コンテンツの録画中であっても、全体の録画終了を待たずに速度変換処理を行うことができる。また、音声区間及び非音声区間の判別結果が蓄積手段に蓄積されることにより、音声含有率の実測値を算出することができ、より最適な速度比で速度変換を行うことができる。

【００４９】

上記第２０の発明によれば、特定イベント含有率を算出して特定イベント区間の速度比を算出することで、例えば特定イベント音が音楽である場合、音楽番組などのオーディオ信号に対して音楽区間を音声区間及び非音声区間よりも遅い再生速度で再生を行うことができる。その結果、音楽を重視した速度変換処理を行うことができる。

【発明を実施するための最良の形態】

【００５０】
以下、本発明の実施形態について、図面を参照しながら説明する。
【００５１】
（第１の実施形態）
まず、図１を参照して本発明の第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、第１の実施形態に係る音声再生装置の構成例を示すブロック図である。図１において、本音声再生装置は、音声非音声判別部１１、蓄積部１２、音声含有率算出部１３、速度比条件設定部１４、音声区間長算出部１５、速度比決定部１６、及び速度変換部１７で構成される。なお、本実施形態では、速度変換対象となるオーディオ信号をコンテンツ単位で予め蓄積部１２に蓄積し、この蓄積したオーディオ信号を用いて再生速度を変えた再生処理を行う音声再生装置について説明する。また以下の説明において、音声が含まれる区間を音声区間とする。また、音声区間以外の区間、つまり音声を含まない区間を非音声区間とする。
【００５２】
音声非音声判別部１１は、オーディオ信号を入力として、音声区間と非音声区間とを判別する。また音声非音声判別部１１は、この判別結果と共に音声区間の始終端時刻（開始時刻及び終了時刻）を出力する。入力されるオーディオ信号は、ＣＤ、ＤＶＤ、メモリ、又はハードディスクなどに記録されたオーディオ信号である。なお、オーディオ信号は、インターネットなどの通信回線を介して配信されたオーディオ信号や放送により受信したオーディオ信号などであってもよい。また、オーディオ信号は音声合成などその場で生成したものや、マイクで収録したもの、電話などの通信機器を通じて出力されるものでもよい。ここで、音声区間及び非音声区間を判別する方法としては、例えばオーディオ信号のパワーを算出し、閾値により判別を行う方法が挙げられる。また例えば「Ｃｅｐｓｔｒｕｍ　Ｆｌｕｘを用いた音声と音楽のセグメンテーション」＜ＳＰ２０００－１、内田貴之、山下昌毅、杉山雅英による，信学技報，ＳＰ２０００－１７＞に記載されるように、ケプストラムの変化度合いを計測して判別を行う方法もある。ケプストラムの変化度合いを計測する方法では、ＢＧＭが重畳した音声であっても判別が可能である。
【００５３】
蓄積部１２は、ハードディスク、ＤＶＤ、又はメモリ媒体（例えばＳＤカード）などの読み書き可能な記録媒体で構成される。蓄積部１２には、音声非音声判別部１１に入力されるのと同じオーディオ信号がコンテンツ単位で蓄積される。また蓄積部１２には、音声非音声判別部１１から出力された、判別結果と音声区間の始終端時刻とが蓄積される。ここで例えば、ＴＶ放送を録画する場合を考える。この場合、ＴＶ放送を構成するオーディオ信号及びビデオ信号は蓄積部１２に蓄積される。またこの蓄積と共に音声非音声判別部１１において判別処理が行われ、判別結果や音声区間の始終端時刻が蓄積部１２に蓄積される。蓄積部１２には、コンテンツ１つに対して、オーディオ信号、ビデオ信号、判別結果、及び音声区間の始終端時刻が対応付けされて蓄積される。なお、オーディオ信号及びビデオ信号のフォーマットは、どのようなフォーマットであってもかまわない。
【００５４】
音声含有率算出部１３は、コンテンツのオーディオ信号に含まれる音声区間の比率を示す音声含有率を算出する。具体的には、音声含有率算出部１３は、蓄積部１２に蓄積された各コンテンツに対して、それぞれに対応する判別結果や音声区間の始終端時刻を用いて音声含有率を算出する。音声含有率は、所定時間のオーディオ信号に含まれる音声区間長の和を当該所定時間で除算したものである。本実施形態では、コンテンツのオーディオ信号に含まれる音声区間長の和をコンテンツ長で除算したものを音声含有率とする。ここでコンテンツとは、速度変換を行う一番組全体を意味する。したがって、コンテンツ長は通常、番組長に等しく、３０分や１時間といったものが多い。なお、ユーザが番組の一部を速度変換対象として指定した場合、その一部をコンテンツとしてもよい。
【００５５】
音声含有率は、コンテンツによって異なる。例えば図２に示すようにコンテンツをジャ

ンル別に見た場合、音声含有率はジャンルによって異なることがわかる。図２は、ジャンル別の音声含有率を示した図である。図２において横軸はジャンルを示し、縦軸は音声含有率を示している。また図２に示す音声含有率は、同一週に放送された番組のうち、ジャンル別の視聴率の上位６位までの番組を抽出して、抽出した番組ごとの音声含有率をジャンル別に集計して平均化したものである。ニュースの音声含有率は、約６０％であり、５ジャンルの中で最も高い値となっている。スポーツや音楽の音声含有率は、約４０％となり、ニュースと比べて２０％近い開きがある。また同じジャンルにおいても、音声含有率には図３に示すような多少のばらつきが存在する。図３は、各ジャンルの音声含有率の平均と標準偏差とを示した図である。ドラマやアニメでは標準偏差が１６．２となり、他のジャンルに比べて高くなっている。このように音声含有率は、コンテンツによって異なる。

【００５６】

したがって、音声含有率を考慮しない従来技術では、上述したように、音声含有率の高いニュース番組などで目標時間からの遅延時間が長くなる。その結果、遅延時間を解消するために部分的に音声区間の高速再生や削除を行い、再生されるオーディオ信号が聞き取り難くなるという問題があった。これに対し、本実施形態では、コンテンツの音声含有率を算出する。これにより、目標時間から遅れることなく、コンテンツに応じた最適な音声及び非音声区間の速度比の算出が可能となり、部分的に偏ることなく聞き取り易い再生を実現することができる。なお、音声含有率を用いた速度比の算出方法については、後述にて詳述する。

【００５７】

速度比条件設定部１４は、音声含有率及び目標圧伸比を入力とし、音声区間の平均速度比、非音声区間の平均速度比、及び音声区間の終端速度比を算出し、これらを速度比条件として設定する。

【００５８】

圧伸比とは、速度変換処理後の再生時間長を速度変換処理前の再生時間長で除算したものである。等倍速の再生では、圧伸比は１となる。２倍速の再生では、圧伸比は０．５となる。圧伸比が０から１までの値をとるとき、再生時間長は圧縮され、等倍速よりも速い速度で再生される。圧伸比が１より大きな値をとるとき、再生時間長は伸張され、等倍速よりも遅い速度で再生される。また目標圧伸比とは、速度変換を行いたいコンテンツの再生時間長をどれぐらい圧縮もしくは伸張するかを示したものである。目標圧伸比は、圧縮の場合は０から１までの値をとり、伸張の場合は１以上の値をとる。目標圧伸比は、ユーザによって入力されてもよいし、予め装置に設定されていてもよい。また、ユーザが目標圧伸比を直接入力しなくてもよい。この場合、コンテンツ再生の目標時間を入力する。ユーザが目標時間を入力した場合、目標時間を速度変換処理前の再生時間長で除算することで、目標圧伸比を得ることができる。また速度比とは、等倍速に対する速度の比率を示したものである。速度比は、圧伸比の逆数で表される。また音声区間の終端速度比とは、音声区間の終端時刻における速度比を意味する。

【００５９】

次に、音声及び非音声区間の平均速度比を算出する方法について説明する。速度比条件設定部１４は、予め設定された速度比算出分布を用いて平均速度比を算出する。速度比算出分布とは、音声含有率及び目標圧伸比に応じて、どの算出パターンで平均速度比を算出するかを示した分布である。換言すれば、速度比算出分布は、音声含有率と、目標圧伸比と、音声及び非音声区間の平均速度比を算出する方法との対応を示した対応情報である。

【００６０】

以下、算出パターンについて説明する。音声及び非音声区間の平均速度比は、目標圧伸比を達成するように算出される。具体的には、式（１）を満たすように算出される。

【数１】

$$\frac{S}{V_{m1}} + \frac{1-S}{V_{m2}} = E \quad \cdots (1)$$

なお、Ｓは音声含有率、Ｖｍ１は音声区間の平均速度比、Ｖｍ２は非音声区間の平均速度比、Ｅは目標圧伸比を示す。算出パターンとしては、図４に示すように５種類の算出パターンａ～ｅが考えられる。図４は、５種類の算出パターンを示した図である。算出パターンａ～ｅの条件は、以下のようになる。

ａ：非音声区間の平均速度比Ｖｍ２＝Ａｎ（固定値）として、与えられる音声含有率Ｓと目標圧伸比Ｅから式（１）を満たすように、音声区間の平均速度比Ｖｍ１を算出する。但し、Ｖｍ１≦Ａｎを算出条件とする。

ｂ：音声区間の平均速度比Ｖｍ１＝Ｂｓ（固定値）として、与えられる音声含有率Ｓと目標圧伸比Ｅから式（１）を満たすように、非音声区間の平均速度比Ｖｍ２を算出する。但し、Ｖｍ２≧Ｂｓを算出条件とする。

ｃ：音声及び非音声区間の平均速度比をＶｍ１＝Ｖｍ２として、与えられる音声含有率Ｓと目標圧伸比Ｅから式（１）を満たすように、音声及び非音声区間の平均速度比Ｖｍ１及びＶｍ２を算出する。

ｄ：非音声区間の平均速度比Ｖｍ２＝Ｄｎ（固定値）として、与えられる音声含有率Ｓと目標圧伸比Ｅから式（１）を満たすように、音声区間の平均速度比Ｖｍ１を算出する。但し、Ｖｍ１≧Ｄｎを算出条件とする。

ｅ：音声区間の平均速度比Ｖｍ１＝Ｅｓ（固定値）として、与えられる音声含有率Ｓと目標圧伸比Ｅから式（１）を満たすように、非音声区間の平均速度比Ｖｍ２を算出する。但し、Ｖｍ２≦Ｅｓを算出条件とする。

【００６１】

このように、平均速度比の算出パターンが異なれば、同じ音声含有率及び目標圧伸比であっても、音声区間の平均速度比と非音声区間の平均速度比の組み合わせは異なることとなる。そこで、音声含有率及び目標圧伸比に応じて、どの算出パターンを選択するか速度比算出分布を用いて決定する。以下、速度比算出分布について説明する。

【００６２】

図５に速度比算出分布の一例を示す。図５において、縦軸は音声含有率、横軸は目標圧伸比を示しており、算出パターンａ～ｃの領域の分布が示されている。ここで、速度比算出分布は、上述した算出パターンから所定の算出パターンを選択し、選択した算出パターンに対して上述した条件を満足しつつ、音声及び非音声区間の平均速度比の取り得る値を設定することで得られる。図５に示す速度比算出分布では、上述した算出パターンのうち算出パターンａ～ｃが選択されている。算出パターンａでは、非音声区間の平均速度比が最大値であるＡｎ＝４、音声区間の平均速度比Ｖｍ１の取り得る値が１．３≦Ｖｍ１≦２と設定されている。この取り得る値は、算出パターンａの算出条件（Ｖｍ１≦Ａｎ）を満足している。算出パターンｂでは、音声区間の平均速度比がＢｓ＝１．３、非音声区間の平均速度比Ｖｍ２の取り得る値が１．３≦Ｖｍ２≦４と設定されている。この取り得る値は、算出パターンｂの算出条件（Ｖｍ２≧Ｂｓ）を満足している。算出パターンｃでは、音声及び非音声区間の平均速度比をＶｍ１＝Ｖｍ２（１≦Ｖｍ１≦１．３）と設定されている。このように算出パターンを選択し、音声及び非音声区間の平均速度比の取り得る値を設定することで、図５の速度比算出分布を得ることができる。音声含有率と目標圧伸比が算出パターンａの領域内にある場合、音声及び非音声区間の平均速度比は、算出パターンａで算出される。算出パターンｂ、ｃについても同様である。このように、音声含有率及び目標圧伸比に応じてどの算出パターンで算出するかが、速度比算出分布によって決まることとなる。なお、図５の一番左側にある処理不可の領域は、音声含有率に対して目標圧伸比が極端に小さく、音声及び非音声区間の平均速度比をユーザが聞き取り可能な範囲で最大にしても、目標圧伸比を達成できない領域である。

【００６３】
なお、図５の速度比算出分布では、音声含有率が高いほど、目標圧伸比に対して算出パターンａによって算出される割合が高くなる。図５の算出パターンａでは、非音声区間の平均速度比が最大値（Ａｎ＝４）に設定されている。これにより、目標圧伸比を達成する上で音声区間の平均速度比Ｖｍ１を可能な限り遅くすることができる。
【００６４】
また、図５の速度比算出分布では、音声含有率が低いほど、目標圧伸比に対して算出パターンｂによって算出される割合が高くなる。図５の算出パターンｂでは、音声区間の平均速度比がＢｓ＝１．３（固定値）に設定されている。これにより、目標圧伸比を達成する上で非音声区間の平均速度比Ｖｍ２を可能な限り遅くすることができる。このように、図５の速度比算出分布は、音声含有率の大きさに応じて音声含有率と算出方法との対応が異なるものとなる。
【００６５】
また、目標圧伸比が大きいとき、算出パターンｃが選択される。つまり、音声と非音声の平均速度比を等しくしている。目標圧伸比が大きいときは、音声と非音声が同じ平均速度比である方が、より自然に再生することができる。このように、図５の速度比算出分布は、目標圧伸比の大きさに応じて目標圧伸比と算出方法とが異なるものとなる。
【００６６】
また、図５に示す速度比算出分布では、音声含有率と目標圧伸比で領域が一意に定まるように、音声及び非音声区間の平均速度比の取り得る値が設定されている。つまり、音声及び非音声区間の平均速度比の取り得る値は、隣り合う算出パターン間の境界で平均速度比の値が連続するように設定されている。具体的には、算出パターンａでは音声区間の平均速度比Ｖｍ１の最下限が１．３であり、隣り合う算出パターンｂの音声区間の平均速度比Ｂｓが１．３である。これにより、算出パターンａ及びｂの境界において平均速度比の値が連続することとなる。また、算出パターンｂでは音声区間の平均速度比Ｂｓが１．３であり、隣り合う算出パターンｃでは音声区間の平均速度比Ｖｍ１の最上限が１．３である。これにより、算出パターンｂ及びｃの境界において平均速度比の値が連続することとなる。
【００６７】
なお、目標圧伸比が大きくなるにつれて音声及び非音声区間の平均速度比がどのように連続して変化するかという観点から説明すると、次のようになる。目標圧伸比が処理不可の領域内の値をとるとき、音声及び非音声区間の平均速度比は算出されない。算出パターンａの領域内の値をとるとき、目標圧伸比が大きくなるにつれて、音声区間の平均速度比は２から１．３まで小さくなる。このとき、非音声区間の平均速度比は、４で一定である。算出パターンｂの領域内の値をとるとき、音声区間の平均速度比は１．３で一定となり、非音声区間の平均速度比は目標圧伸比が大きくなるにつれて４から１．３まで小さくなる。算出パターンｃの領域内の値をとるとき、音声及び非音声区間の平均速度比は、共に同じ値となりながら、１．３から１まで小さくなる。
【００６８】
このように、速度比算出分布が、算出パターンの切り替わる境界の平均速度比が連続値となるように設定されることで、平均速度比が不連続な値をとる際に急激な速度変換が起こり、違和感が生じるという問題を回避することができる。
【００６９】
図６は、音声含有率が０．５のときの目標圧伸比、音声区間の平均速度比、非音声区間の平均速度比を示している。上述した図５に示されるように、目標圧伸比が０．３７５から０．５１０までの値をとるとき、算出パターンａが選択される。目標圧伸比が０．５０１から０．７６９までの値をとるとき、算出パターンｂが選択される。目標圧伸比が０．７６９から１までの値をとるとき、算出パターンｃが選択される。ここで、図５に示した各算出パターンには、上述したように、平均速度比の取り得る値が設定されている。従って、上述した算出パターン及び式（１）により、図６に示すような平均速度比が算出され

る。
【００７０】
目標圧伸比が０．１及び０．３の値をとるとき、図５に示す速度比算出分布からも分かるように、処理不可の領域となるので、音声及び非音声区間の速度比は算出されない。目標圧伸比が０．４及び０．５の値をとるとき、共に算出パターンａによって算出される。なお、算出パターンａによって算出される場合、目標圧伸比が増加するにつれて音声区間の平均速度比が小さくなっていることが分かる。目標圧伸比が０．６及び０．７の値をとるとき、共に算出パターンｂによって算出される。なお、算出パターンｂによって算出される場合、目標圧伸比が増加するにつれて非音声区間の平均速度比が小さくなっていることが分かる。目標圧伸比が０．９及び１．０の値をとるとき、音声及び非音声区間の平均速度比が等しくなり、目標圧伸比が増加するにつれて音声及び非音声区間の平均速度比が小さくなっていることが分かる。
【００７１】
次に、音声区間の終端速度比を算出方法について説明する。速度比条件設定部１４は、算出した音声区間の平均速度比から、音声区間の終端速度比を算出する。図７は、音声及び非音声区間の速度比変化を示した模式図である。音声区間１の区間長は、音声区間２の区間長よりも短くなっている。縦軸は変換速度比であり、横軸は経過時間である。変換速度比とは、速度変換部１７の速度変換処理に用いられる速度比を示しており、音声及び非音声区間をそれぞれ細分化した各区間における速度比によって示される。変換速度比の決定方法については、後述にて説明する。図７に示すように、音声区間長が異なっていても、音声区間の終端時刻での速度比は等しくなっている。速度比条件設定部１４は、この終端時刻の速度比を音声区間の終端速度比として算出している。図７からも明らかなとおり、音声区間は終端速度比よりも遅い速度比が設定されている。音声区間の終端速度比Ｖｅｎｄは音声区間の平均速度比Ｖｍ１にαを加算したものとする。つまり、終端速度比Ｖｅｎｄ＝Ｖｍ１＋αとする。なお、聴取実験により、αを０．２とし、Ｖｅｎｄは２．０を超えないものが好ましいことが分かった。なお、図７では、非音声区間の速度比は平均速度比Ｖｍ２ｅｎｄで一定である。
【００７２】
以上のように、速度比条件設定部１４は、音声含有率及び目標圧伸比を入力とし、音声区間の平均速度比、非音声区間の平均速度比、及び音声区間の終端速度比を算出し、これらを速度比条件として設定する。
【００７３】
音声区間長算出部１５は、音声区間の始終端時刻を入力とし、音声区間長を算出する。速度比決定部１６は、速度比条件設定部１４で設定された速度比条件と音声区間長とに基づき、音声及び非音声区間の変換速度比を決定する。ここで、変換速度比とは、上述したように、速度変換部１７の速度変換処理に用いられる速度比を示しており、音声及び非音声区間をそれぞれ細分化した各区間における速度比によって示される。ただし、音声区間中の速度比を一定にする場合や、一定時間ごとに速度比を切り替える場合は音声区間長を必ずしも算出する必要はないため、音声区間長算出部１５を設けなくてもよい。たとえ、音声区間長算出部１５を設けなかったとしても、音声含有率によって音声区間の平均速度比を設定しているため、従来の方法よりも聞き易くなる。これに対し、音声区間長算出部１５を設けることによって、音声区間長を算出し、音声区間の細分化された各区間に対して速度比を設定することで、更に聞き易くなる効果がある。
【００７４】
速度変換部１７は、オーディオ信号を入力とし、速度比決定部１６で決定された変換速度比に従って速度変換を行う。速度変換の方法としては、例えば「高品質音声速度変換方式のＤＳＰによる実現」＜鈴木，三崎，電子情報通信学会　音声研究会資料　ＳＰ９０－３４、（１９９０．８．２３）＞、特許第３１８９５８７号などに記載された公知の方法を用いるとする。このような方法により、１倍速以下の遅い速度比での再生や、１倍速以上の速い速度比での再生が可能となる。また、速度変換の方法はこれに限らず、音を合成

したり、区間の削除や挿入などを行ったり、速度比決定部１６で決定された変換速度比を満たすような処理を行っているものであれば方法は問わない。例えば、変換速度比が０．５である場合を仮定すると、ある入力区間に対して出力再生時間が２倍となればよく、音を引き伸ばしたり、無音区間を追加したり、新たに音を合成してもよい。このように、速度変換部１７はある区間に対する入力と出力との関係が対応付けられており、変換速度比を満たすような処理を行っているものであれば、速度変換の方法として含まれる。

【００７５】

以下、図８を参照して、第１の実施形態に係る音声再生装置の処理について説明する。図８は、第１の実施形態に係る音声再生装置の処理の流れを示すフローチャートである。

【００７６】

まず、ユーザが入力装置（図示なし）においてコンテンツを録画する指示をしたとき、当該コンテンツのオーディオ信号及びビデオ信号が蓄積部１２に蓄積される。このとき、音声非音声判別部１１はそのコンテンツのオーディオ信号について音声区間と非音声区間とを判別する（ステップＳ１０１）。なお、ステップＳ１０１において判別された判別結果と音声区間の始終端時刻についても、蓄積部１２に蓄積される。

【００７７】

ステップＳ１０１の次に、入力装置において、ユーザが所望のコンテンツを再生する指示をしたか否かが判断される（ステップＳ１０２）。ユーザの指示があった場合（ステップＳ１０２でＹｅｓ）、音声含有率算出部１３は、指示されたコンテンツの音声含有率を算出する（ステップＳ１０３）。

【００７８】

ステップＳ１０３の次に、ユーザが入力装置（図示なし）において目標圧伸比を設定する（ステップＳ１０４）。速度比条件設定部１４は、ステップＳ１０３で算出された音声含有率と、予め設定された速度比算出分布とから、ステップＳ１０４で設定された目標圧伸比が処理不可の領域内にあるか否かを判断する（ステップＳ１０５）。処理不可の領域内に目標圧伸比が設定された場合（ステップＳ１０５でＮｏ）、処理はステップＳ１０４に戻る。ステップＳ１０４において、速度比条件設定部１４は、目標圧伸比を処理可能な値に再設定する。図５の場合、音声含有率が０．５のとき処理可能な最小の目標圧伸比は、０．３７５となる。したがって速度比条件設定部１４は、最も近い領域境界の値０．３７５を目標圧伸比として再設定する。なお、速度比条件設定部１４が自動で再設定するのではなく、目標圧伸比の入力を再度ユーザに求めるようにしてもよい。

【００７９】

ステップＳ１０５の次に、速度比条件設定部１４は、ステップＳ１０３で算出された音声含有率、ステップＳ１０４及びＳ１０５で設定された目標圧伸比に基づいて、音声区間の平均速度比、非音声区間の平均速度比、及び音声区間の終端速度比を算出し、速度比条件を設定する（ステップＳ１０６）。なお、速度比条件の算出方法については、上述したとおりである。

【００８０】

ステップＳ１０６の次に、音声区間長算出部１５は、音声区間の始終端時刻を入力とし、音声区間長を算出する（ステップＳ１０７）。音声区間長は、図９及び図１０に示すように、同じコンテンツ内でも長短様々なものが含まれているが、ジャンルによっても大きく異なる。図９は、ニュース番組に含まれる音声区間長とその頻度を示した図である。図１０は、野球番組に含まれる音声区間長とその頻度を示した図である。図９及び図１０において、横軸は音声区間長であり、縦軸は番組中に発生した頻度である。

【００８１】

ここで、上述した従来技術では、音声区間長を考慮せず始端からの経過時間のみで速度比を設定しており、音声区間長が長いものでは経過時間に伴って速度比が段々速くなり、聞きにくくなるという課題があった。これに対し、本実施形態では、音声区間長を考慮することで、図７に示したように、音声区間長の長短に関わらず、音声区間の終端での速度比が等しくなるように速度比を決定することができる。これにより、音声区間の速度比が

始端から徐々に速くなるが、終端が聞き取り可能な速度比までしか速くならないため、従来技術に比べ、聞き易さが大きく改善した。

【００８２】

ステップＳ１０７の次に、速度比決定部１６は、蓄積部１２に蓄積された音声区間の始終端時刻を参照して、コンテンツの始端から順に所定の単位時間毎に音声区間であるか否かを判断する（ステップＳ１０８）。音声区間と判断した場合、速度比決定部１６は、音声区間の始終端時刻と、ステップＳ１０７で算出された音声区間長とに基づき、音声区間における経過割合を算出する（ステップＳ１０９）。音声区間の経過割合とは、音声区間の始端を０、終端を１として、始端からの経過時間を音声区間長で除算したものである。

【００８３】

ステップＳ１０９の次に、速度比決定部１６は、音声区間の経過割合から、音声区間の変換速度比を決定する（ステップＳ１１０）。以下、ステップＳ１１０の具体的な処理例について説明する。変換速度比の算出処理の一例としては、音声区間の平均圧伸比に対する圧伸比変化量の和が０になるように、変換速度比を算出する方法が挙げられる。図１１は、音声区間の圧伸比の変化を示した図である。図１１において、ｘは経過割合、ｖｘは経過割合がｘのときの変換圧伸比、ｖｓは始端圧伸比、ｖｅは終端圧伸比、ｖｍ１は平均圧伸比とする。ここで、始端圧伸比ｖｓと終端圧伸比ｖｅとを結ぶ圧伸比の変化カーブは、式（２）で表現される。

【数２】

$$v_x = (v_e - v_s)x + v_s \qquad \cdots (2)$$

平均圧伸比ｖｍ１は、音声区間の平均速度比Ｖｍ１の逆数である。終端圧伸比ｖｅは、終端速度比Ｖｅｎｄの逆数である。ここで、圧伸比変化量は、音声区間の平均圧伸比ｖｍ１を０と想定したとき、ｖｍ１に対して増減する量（図１１の網掛け部分の面積）を意味する。したがって、この量の和が０となるためには、図１１に示したように、ｘ＝０．５のときに変換圧伸比ｖｘが平均圧伸比ｖｍ１となるようにすればよい。平均圧伸比ｖｍ１は、音声区間の平均速度比Ｖｍ１から求まる値であり、終端圧伸比ｖｅは、終端速度比Ｖｅｎｄから求まる値である。したがって、式（３）を満たすように始端圧伸比ｖｓを設定すればよいこととなる。

【数３】

$$v_{m1} = \frac{1}{2}(v_e - v_s) + v_s \qquad \cdots (3)$$

なお、経過割合がｘのときの変換圧伸比ｖｘ、始端圧伸比ｖｓ、終端圧伸比ｖｅ、平均圧伸比ｖｍ１を速度比で表すと、式（４）のようになる。ここで、Ｖｘは経過割合がｘのときの変換速度比、Ｖｓは始端速度比、Ｖｅｎｄは上述した終端速度比、Ｖｍ１は上述した平均速度比を示す。

【数４】

$$v_{m1} = \frac{1}{V_{m1}}, \quad v_e = \frac{1}{V_{end}}, \quad v_s = \frac{1}{V_s}, \quad v_x = \frac{1}{V_x} \quad \cdots (4)$$

そして、式（２）に式（３）および（４）を代入すると、式（５）が得られる。

【数５】

$$V_x = \frac{V_{end} V_s}{(V_s - V_{end})x + V_{end}} \quad \cdots (5)$$

なお、速度変換後の音声区間長は、平均速度比Ｖｍ１で一様に変換した時間長と等しくなることから、式（６）が成り立つ。

【数６】

$$V_s = 2V_{m1} - V_{end} \quad \cdots (6)$$

このようにステップＳ１１０において、速度比決定部１６は、式（５）に音声区間の経過割合ｘを代入することで、音声区間の変換速度比Ｖｘを決定することができる。このステップＳ１１０で算出した音声区間の変換速度比は、上述した図７のような変化となる。つまり、音声区間の冒頭部分を遅くし、終端に向かって徐々に速めていくように、変換速度比を音声区間長に応じて変化させることができる。

【００８４】

ステップＳ１０８において非音声区間と判断した場合、速度比決定部１６は、非音声区間の始端から終端まで、速度比条件設定部１４で設定された非音声区間の平均速度比を変換速度比として決定する。つまり、図７に示したように、速度比決定部１６は、平均速度比で一定となるように、非音声区間の始端から終端までの変換速度比を決定する。

【００８５】

ステップＳ１１０及びＳ１１１の次に、速度比決定部１６は、コンテンツの終端まで変換速度比を算出したか否かを判断する（ステップＳ１１２）。終端ではないとき、処理はステップＳ１０８へ戻る。このように、コンテンツの終端までの変換速度比が算出されるまで、速度比決定部１６においてステップＳ１０８～Ｓ１１２までの処理が繰り返される。ステップＳ１１２においてコンテンツの終端まで変換速度比が算出されたと判断された場合、速度変換部１７において変換速度比に従ってオーディオ信号の速度変換が行われ、速度変換後のオーディオ信号の再生が開始される（ステップＳ１１３）。入力装置（図示なし）が本装置の処理を終了するか否かの指示を受け付ける（ステップＳ１１４）。ユーザが他のコンテンツについて速度変換処理を行う場合（ステップＳ１１４でＮｏ）、処理はステップＳ１０２へ戻る。

【００８６】

以上のように、本実施形態に係る音声再生装置によれば、コンテンツの音声含有率を算出することにより、コンテンツに応じた速度比条件を設定することができる。これにより、目標圧伸比、つまり目標時間を達成しつつも、音声区間及び非音声区間の速度比をコンテンツに応じた最適な速度比にそれぞれ設定することができる。その結果、どのようなコンテンツのオーディオ信号が入力されても、聞き取り易い速度で再生することが可能となり、再生内容の不連続性や情報の欠落による不快感などを低減させた再生を行うことができる。

【００８７】

また本実施形態に係る音声再生装置によれば、速度比決定部１６において図１１に示すように圧伸比の変化を示す関数として、１次関数を用いるとした。つまり、本装置に入力されたオーディオ信号に対して、一次直線で速度比を設定している。ここで、日本語はモーラリズムの言語と言われており、個々のモーラが同じ長さになるように話す傾向がある。モーラは言葉を話すときの長さの単位であり、日本語では俳句や短歌で数える拍に相当する。「かな」でいえば、一文字に相当している。このモーラ毎に速度比を変化させることが望ましいが、入力されるオーディオ信号に対して一次直線で速度比の算出を行っているので、モーラ毎に速度比を変化させなくても、十分に自然な再生を実現することができ

る。さらに、音声区間の始端から終端までの速度比は、一次関数によって、細かく切り替えられている。これにより、知覚される時間よりも短い間隔で速度を変化させることとなり、違和感が少ない自然な再生を提供することができる。

【００８８】

また本実施形態に係る音声再生装置によれば、音声区間の始端から終端まで音声区間長に応じて速度比を設定している。これにより、音声区間の終端時刻において予め設定した終端速度比よりも速い速度比になることなく、音声区間の終端時刻付近において速度比が速くなりすぎて聞き取り難くなることを防ぐことができる。

【００８９】

なお、上述では図１１に示したように、圧伸比の変化を一次関数によって表すようにしたが、他の関数によって表されても構わない。例えば、上に凸または下に凸の指数関数であってもよい。また例えば、予め用いることができる速度比が限られている場合は、２段階や数段階の速度比で変換をおこなっても、音声区間長の経過割合に応じた速度変換を行うことで、不自然さを低減させた再生を提供することができる。図１２は、２段階の変換速度比を算出した場合を示す図である。図１２において、より好ましくは、音声区間全体に対して、最初の変換速度比が始端から２～３割の範囲を占めるようにする。これにより、より自然な再生を実現することが聴取実験で明らかとなった。また例えば、音声区間の速度比を非音声区間と同様に一定の速度比としてもよい。この場合であっても、音声含有率により、適切な速度比が設定されるため、従来技術のような非音声区間の削除や極端な高速化が行われずに済み、聞き易い再生を提供することができる。

【００９０】

なお、上述では、速度比条件設定部１４は、図５に示した速度比算出分布を用いるとしたが、これに限定されない。ユーザ自身が、所望の算出パターンを選択して音声及び非音声区間の平均速度比の取り得る値を所望の値に設定し、速度比算出分布を作成するようにしてもよい。つまり、速度比条件設定部１４が用いる速度比算出分布は、予め設定されているものに限らず、ユーザによって設定されるものであってもよい。例えば、図５では音声区間の平均速度比が２．０までとり得る。しかし、高齢者や語学学習者では２．０よりももっと遅い平均速度比での聴き取りを希望する場合もある。その際に、ユーザが望む平均速度比を超えないように、音声及び非音声区間の平均速度比の取り得る値を設定することで、ユーザの聴き取り能力に応じた再生処理が可能となる。また、高齢者や語学学習者では通常の再生速度よりさらに遅くして聞きたい場合が存在する。音声区間は通常の平均速度比１．０よりも遅い速度にし、非音声区間を通常の平均速度比１．０より高速化して通常の再生時間と同じ時間内で収めたい、あるいはもっと短い時間で視聴したいといった要望に答えるためにも、速度比条件設定部１４の速度比算出分布は用途に応じて切り替えることを可能にしている。

【００９１】

また、図５に示した速度比算出分布は、ジャンル毎に予め用意されていてもよい。この場合、ＥＰＧ等のジャンル情報やユーザの指示によって、いずれの速度比算出分布を用いるかを選択する。ここで音声含有率以外にも、画像の動きの激しさ等はジャンルによって異なる。例えば、ドキュメンタリーなどの静止画像が多いジャンルでは、非音声区間を高速で再生しても、画像の高速化による情報の欠落は少ない。また、非音声区間を高速で再生できるので、音声区間を１倍速に近づけることができる。その結果、番組内容を理解しやすい再生を行うことができる。ここで、ドキュメンタリーなどの静止画像が多いジャンルについての速度比算出分布の例を図１３に示す。図１３に示すように、速度比算出分布は、算出パターンａ及びｂの領域で構成される。このうち、算出パターンａでは、非音声区間の平均速度比が最大値であるＡｎ＝４、音声区間の平均速度比Ｖｍ１の取り得る値が１≦Ｖｍ１≦２と設定されている。算出パターンｂでは、音声区間の平均速度比がＢｓ＝１、非音声区間の平均速度比Ｖｍ２の取り得る値が１≦Ｖｍ２≦４と設定されている。

【００９２】

図１４は、図１３に示す速度比算出分布において、音声含有率が０．５のときの目標圧

伸比、音声区間の平均速度比、非音声区間の平均速度比を示している。目標圧伸比が０．１及び０．３の値をとるとき、図１３に示す速度比算出分布からも分かるように、処理不可の領域となるので、音声及び非音声区間の速度比は算出されない。目標圧伸比が０．４、０．５、及び０．６の値をとるとき、共に算出パターンａによって算出される。目標圧伸比が０．７、０．９、及び１の値をとるとき、共に算出パターンｂによって算出される。図１４から分かるように、図１３に示す速度比算出分布を用いた場合、算出パターンが２つのパターンで構成されるので、音声及び非音声の平均速度比の差が大きくなる。換言すれば、非音声区間を高速化し、音声区間を１倍速にすることができ、番組内容をより理解し易い再生を実現することができることを意味する。

【００９３】

また、例えばスポーツなど動きの激しいシーンが多いジャンルでは、音声と非音声の平均速度比に大きな差をつけないほうがよい。なぜならば、動きの激しいシーンが多いジャンルは、動きの少ないジャンルに比べ、番組の内容理解に対して音声以外の部分が与える影響が大きいため、非音声区間の聞き取り易さや見易さを向上させる必要があるからである。ここで、スポーツなど動きの激しいシーンが多いジャンルについての速度比算出分布の例を図１５に示す。図１５に示すように、速度比算出分布は、２つの算出パターンａ、２つの算出パターンｂ、及び算出パターンｃの領域で構成される。このうち、一番左側の算出パターンａでは、非音声区間の平均速度比がＡｎ＝３、音声区間の平均速度比Ｖｍ１の取り得る値が１．８≦Ｖｍ１≦２．５と設定されている。この算出パターンａと隣り合う算出パターンｂでは、音声区間の平均速度比がＢｓ＝１．８、非音声区間の平均速度比Ｖｍ２の取り得る値が２．５≦Ｖｍ２≦３と設定されている。この算出パターンｂと隣り合う算出パターンａでは、非音声区間の平均速度比がＡｎ＝２．５、音声区間の平均速度比Ｖｍ１の取り得る値が１．５≦Ｖｍ１≦１．８と設定されている。この算出パターンａと隣り合う算出パターンｂでは、音声区間の平均速度比がＢｓ＝１．５、非音声区間の平均速度比Ｖｍ２の取り得る値が１．５≦Ｖｍ２≦２．５と設定されている。この算出パターンｂと隣り合う算出パターンｃでは、音声及び非音声区間の平均速度比をＶｍ１＝Ｖｍ２（１≦Ｖｍ１≦１．５）と設定されている。

【００９４】

図１６は、図１５に示す速度比算出分布において、音声含有率が０．５のときの目標圧伸比、音声区間の平均速度比、非音声区間の平均速度比を示している。目標圧伸比が０．１及び０．３の値をとるとき、図１５に示す速度比算出分布からも分かるように、処理不可の領域となるので、音声及び非音声区間の速度比は算出されない。目標圧伸比が０．４及び０．５の値をとるとき、共に算出パターンａによって算出される。目標圧伸比が０．６の値をとるとき、算出パターンｂによって算出される。目標圧伸比が０．７、０．９、及び１の値をとるとき、算出パターンｃによって算出される。図１５から分かるように、目標圧伸比に対して算出パターンは多数切り替わっている。これにより、図１５に示す速度比算出分布を用いた場合、音声と非音声の平均速度比に大きな差が生じない。その結果、非音声区間の聞き取り易さ及び見易さが向上する。また、図６で示した速度比算出分布よりも、非音声区間の速度比を若干遅めに設定している。これにより、非音声区間において生じる動きが激しいシーンが多いジャンルに対して、非音声区間の聞き取り易さ及び見易さをさらに向上させることができる。このように、速度比算出分布をジャンル毎に準備しておくことで、より的確な速度変換処理が可能になる。

【００９５】

なお、ジャンル毎だけではなく、動きの激しさなどを示す画像情報や、音響的な特徴に応じた速度比算出分布が予め用意されていてもよい。このような速度比算出分布は、例えば、音楽やある特定の人物の音声などユーザが着目したい音に対して個別に速度を制御したい場合に、有効である。

【００９６】

また、上述では、速度比算出分布を構成する領域として、算出パターンａ～ｃを用いて場合について説明したが、目的に合わせて算出パターンｄ及びｅを用いてもよい。音楽番

組において音楽を重視して再生したい場合、音楽は非音声区間であるため、出来るだけ非音声区間の速度比を遅くするとよい。その代わり、音声区間の速度比を速くする必要がある。したがって、このような音楽番組などのジャンルに対しては、上述した算出パターンｄを用いるのが好適である。これは音楽に限らず、ユーザが着目したい音に対して再生を行う場合に有効である。また、ユーザがコンテンツに含まれる非音声区間をサーチする場合、音声区間を出来るだけ高速で再生することが望まれる。したがって、この場合、上述した算出パターンｅを用いるのが好適である。

【００９７】

また、速度比条件設定部１４は、音声区間長算出部１５で得られた音声区間長から、図９及び図１０に示したような統計的な分布を求め、その統計的な分布に基づく速度比算出分布を用いてもよい。音声区間長とその生起頻度はコンテンツの属性を示している。このため、統計的な分布に基づく速度比算出分布を用いることで、コンテンツの属性に応じた速度変換処理が可能になる。例えば音声含有率が同じであったとしても、音声区間長が短いものが多く音声区間の生起頻度が高いコンテンツや、音声区間長が長いものが多く音声区間の生起頻度が低いコンテンツが存在する。後者のコンテンツでは、一つの音声区間あたりに含まれる情報量が多く、理解にかかるユーザの負荷が高いことが想定される。したがって、このようなコンテンツに対しては、音声区間に対してより重点的に遅い速度比を配分するような速度比算出分布を用いる。このように速度比条件設定部１４は、音声区間長の統計的な分布に基づく速度比算出分布を用いて、速度比条件を設定してもよい。このことは、プライベートコンテンツについて特に有効である。プライベートコンテンツは放送コンテンツと異なり編集等の処理を施していないものが多いため、音声含有率や音声区間長もコンテンツごとにばらつきが大きい。そのため、様々な速度比算出分布を用意することで、プライベートコンテンツなどコンテンツ間で音声区間長や音声含有率のばらつきが大きいものにおいても適切な速度比条件を設定することが可能になる。

【００９８】

また、上述では、目標圧伸比が０から１となる場合についてのみ説明を行ったが、速度制御後の再生時間が通常再生時間と同じかそれよりも長い時間で視聴を行う遅聞きや遅見再生など目標圧伸比１以上の場合についても、同様に速度比算出分布を予め用意しておくことで、速度比条件を設定することが可能である。また、一つの音声区間ごとに音声区間長と同じ長さの非音声区間を設け発音練習を促すような発音練習モードの出力制御も可能となる。例えば、発音練習モードでは、非音声区間の平均速度比は直前の音声区間長によって定まるとすると、音声区間や非音声区間の速度比は以下の式から算出できる。

【数７】

$$V_{m1} = \frac{2*S}{E} \qquad \cdots (7)$$

【数８】

$$V_{m2}(n) = \frac{N(n)*V_{m1}}{M(n)} \quad \cdots (8)$$

なお、Ｓは音声含有率、Ｖｍ１は音声区間の平均速度比、Ｖｍ２は非音声区間の平均速度比、Ｅは目標圧伸比を示す。ｎ番目の音声区間の音声区間長をＭ（ｎ）とし、ｎ番目の音声区間に後続する非音声区間の非音声区間長をＮ（ｎ）とする。音声区間長と同じ長さの非音声区間長を設けるため、音声区間の平均比Ｖｍ１は式（７）のように表せる。また発音練習を行うには、音声区間と同じ長さの非音声区間を必要とする。このため、後続の非音声区間の速度比は、音声区間長に応じて算出する必要があるため、式（８）のようになる。このように、音声区間長と同じ長さの非音声区間を設け、一定の時刻内で再生するという発音練習モードにおいても、音声含有率や音声区間長を利用することで、適切な速度

比を設定することが可能になる。なお、今回は音声区間と同じ長さの非音声区間を設けたが、非音声区間の長さは学習の進み具合などに応じて変更してもよい。このように、語学学習用に新たにコンテンツを作成しなくても、音声含有率と音声区間長の利用によって、発音練習に適した速度で音声を提示することが可能になる。また、学習に費やしたい時間を始めに指定することで、コンテンツ長から圧伸比を算出し、学習時間内におさまるように速度を制御することが可能になる。また、学習のレベルに応じて速度比を変えることも可能となる。

【００９９】

他の使用目的としては、音声の書き起こしを行うときに用いる書き起こしモードが考えられる。この場合も同様に速度比条件設定部１４の速度比算出分布を変えることで対応可能である。音声を書き起こすには書き起こす人の書き込み能力、例えば、紙に記入する場合、一定時間で何文字記入可能かということや、キーボードで打ち込む場合、一定時間に何タイプ可能かなど各ユーザの書き込み能力に応じて、再生速度を変える必要がある。書き込み能力より再生速度が速ければ、すぐに書き込み部分が追い越され、書き込み部分より先の部分が再生されてしまう。そのため、一時停止や巻き戻しなどの操作が必要となる。また、そのような再生制御の操作は書き込み処理を中断させるため、二重に時間を無駄に消費させることになる。書き込み能力より再生速度が遅ければ再生部分に追い越されることはないが、書き込み後も次の音声が始まるまで待ち時間が発生し無駄な時間を消費することに変わりは無い。そこで、音声の書き起こしを行うときはユーザの書き込み能力に応じた速度で再生する必要がある。従来の方法では音声区間長や音声含有率が不明なため、非音声区間も音声区間と同じ速度で再生されたり、速度を遅くした場合どの程度の時間がかかるかは事前にわからなかったりした。今回速度条件設定部１４で以下のような処理をおこなえば、音声区間は聞きやすく非音声区間を省くような再生が可能となる。音声区間の平均速度比は式（９）のように表せる。

【数９】

$$V_{m1} = \frac{Q*S}{Q*E - U*P} \qquad \cdots (9)$$

【数１０】

$$V_{m2} = \infty \qquad \cdots (10)$$

なお、Ｓは音声含有率、Ｖｍ１は音声区間の平均速度比、Ｖｍ２は非音声区間の平均速度比、Ｅは目標圧伸比を示す。コンテンツに含まれる音声区間の総数をＵとし、コンテンツの全長をＱとしている。Ｐは音声区間を再生後に、書き込み行の変更など書き起こし作業用に必要な時間として設けた一定時間である。従って作業によってはＰの値は０でも構わない。非音声区間は再生されないため、非音声区間の速度比Ｖｍ２は式（１０）のようになる。ここで、音声区間の平均速度比Ｖｍ１をユーザの能力に応じた書き込み速度に設定すれば、書き込み作業中に一時停止や早送り、巻き戻しなどの機器操作をすることなく、書き起こし作業が可能となる。このように速度条件設定部１４で音声区間と非音声区間の平均速度比を設定することで、ユーザの書き込み能力に応じた速度比で音声区間のみを再生するような書き起こし作業モードが可能となる。

【０１００】

その他にも、学習開始時には音声区間を遅い速度で再生を行い、学習経過時間に応じて徐々に速度比を早くしながら学習時間の合計は所定の時間になるように、音声や非音声区間の速度を制御するような聴き取り練習用の制御も可能になる。例えば、一回の学習の中で、学習開始時には学習終了時に比べ、遅い速度で音声区間の再生を行い徐々に音声区間の再生速度を上げていくような制御を行うこともできる。また長期的な学習の中で、最初

に学習を開始した時点からの経過時間や前回の操作履歴などから、今回の学習における音声区間の再生速度を制御してもよい。このように音声区間や非音声区間の再生速度を変更したり、音声区間の直前や直後に一時停止や無音やオーディオ区間を付与したり、画面を挿入したりしながらコンテンツ全体の時間調整をすることにも対応可能である。

【０１０１】

（第２の実施形態）

図１７を参照して、本発明の第２の実施形態に係る音声再生装置について説明する。図１７は、第２の実施形態に係る音声再生装置の構成例を示すブロック図である。図１７において、本音声再生装置は、音声非音声判別部１１、音声含有率予測部１８、速度比条件設定部１４、音声区間長予測部１９、速度比決定部１６、速度変換部１７、及び圧伸比算出部２０で構成される。本実施形態は、第１の実施形態に係る音声再生装置に対し、オーディオ信号の再生速度を変えた再生処理をリアルタイムで行う点で異なる。以下、異なる点を中心に説明する。

【０１０２】

音声含有率予測部１８は、音声非音声判別部１１から出力されたフレーム毎の判別結果から、現時点より過去数分の音声含有率を算出する。そして、算出した音声含有率を用いて、現時点より１つ進んだセクションにおける音声含有率を予測する。現時点より過去数分の音声含有率をＸ（ｚ－１）とすると、音声含有率の予測値Ｙ（ｚ）は式（１１）で表現される。Ｙ（ｚ－１）は一つ前のセクションでの音声含有率の予測値である。以下、音声含有率の予測値を予測音声含有率と称す。αは０から１までの値で、シミュレーションにより最適な値を求めた。

【数１１】

$$Y_{(z)} = \alpha Y_{(z-1)} + (1-\alpha) X_{(z-1)} \qquad \cdots (11)$$

式（１１）において、初期値Ｘ（０）＝Ｙ（０）は、一般的なコンテンツの音声含有率の平均値とする。また音声含有率Ｘ（ｚ）が算出されるまでの間予測音声含有率Ｙ（ｚ）はＹ（０）を維持するとする。図１８に、式（１１）で示される予測音声含有率Ｙ（ｚ）を求める方法を模式的に示す。図１８に示すように、現時点より過去数分の音声含有率は、フレーム毎に移動しながら算出される。また、予測音声含有率Ｙ（ｚ）やＹ（ｚ－１）がセクション毎に予測される。図１８に示すように、音声含有率予測部１８は、一つ前のセクションでの予測音声含有率Ｙ（ｚ－１）と、現時点より過去数分の音声含有率Ｘ（ｚ－１）を用いて予測している。

【０１０３】

ここで音声含有率Ｘ（ｚ）の算出時間が短すぎると、音声区間長が長いものであれば、音声含有率が１となってしまう。また、音声と音声の間にある短いポーズ区間のみを抽出してしまい、音声含有率が０に近くなるなど、極端な値をとる可能性がある。また、算出時間が長すぎると、平滑化してしまい、音声含有率の予測に利用できなくなる。そのため、音声含有率Ｘ（ｚ）の算出時間は、コンテンツの音声区間の集まり具合を適度に表す必要があり、実験の結果、１分以上が望ましいことがわかった。したがって、上述では過去数分としている。図１９に音声含有率Ｘ（ｚ）と予測音声含有率Ｙ（ｚ）の算出結果の一例を示す。このグラフは、縦軸に音声含有率を横軸に番組の経過時間を示したものである。また図１９においては、３０分番組のニュース番組について、音声含有率Ｘ（ｚ）と式（１１）によって算出した予測音声含有率Ｙ（ｚ）を図示したものである。図１９に示すように、予測音声含有率Ｙ（ｚ）は、実際の音声含有率Ｘ（ｚ）とほぼ同じ値に推移することが分かる。

【０１０４】

圧伸比算出部２０は、セクション単位で圧伸比を算出する。具体的には、まず現時刻ｔにおける圧伸比を算出する。現時刻ｔにおける圧伸比は、現時刻ｔにおいて速度変換部１

７から出力された出力データ量を速度変換部１７に入力された入力データ量で除算することで求められる。次に圧伸比算出部２０は、現時刻ｔがセクション境界に達したどうかを判断する。セクション境界に達したとき、次のセクションでの速度比条件を設定するため、次のセクションにおける圧伸比（以下、セクション圧伸比と称す）を算出する。セクション圧伸比とは、次のセクションをどのくらいの圧伸比で変換するかを定めたものである。圧伸比算出部２０は、ユーザ又は機器が予め設定した目標圧伸比と、現時刻ｔにおける圧伸比とから、式（１２）を用いて算出する。式（１２）において、Ｒｔは目標圧伸比、Ｒ（ｔ）は現時刻ｔにおける圧伸比、次のセクションの時間長をＴ（ｚ）、Ｒｓ（ｚ）は次のセクションの圧伸比とする。

【数１２】

$$Rs_{(z)} = Rt + (Rt - R_{(t)})\frac{t}{T_{(z)}} \qquad \cdots (12)$$

【０１０５】

速度比条件設定部１４は、音声含有率予測部１８で算出された予測音声含有率と、圧伸比算出部２０で算出されたセクション圧伸比を入力とする。速度比条件設定部１４は、第１の実施形態で説明した速度比条件設定部１４と同様の方法で、セクション毎に速度比条件を設定する。

【０１０６】

ここで、上述した遂次的な処理を行う場合、従来技術では局所的な音声区間の偏りが生じた場合に、その偏りが生じた箇所において再生時間を達成しようと処理するので、非音声区間の極端な削除や高速化が局所的に生じていた。これに対し、本実施形態では、セクション圧伸比がセクション単位で算出される。また速度比条件設定部１４は、セクション単位で、速度比条件を設定する。つまり、速度比条件は、セクション圧伸比によってセクション単位に更新される。これにより、偏りが生じた箇所において再生時間を達成しなくても、次以降のセクションへ持ち越すことができるので、目標時間を達成しつつ、聞き取り易い再生を実現した逐次的な処理を行うことができる。

【０１０７】

なお、予測音声含有率をそのまま利用すると、予測音声含有率の増減と平均速度比の増減が直結してしまう。予測音声含有率が高い部分で平均速度比が速くなることは聞こえに影響を与える可能性がある。なぜならば、例えば発話速度が同じと仮定すると、音声含有率が高いほど、情報量が多いため、文章を理解するのが難しいからである。そこで、予測音声含有率が増加すると平均速度比を下げ、予測音声含有率が減少すると平均速度比が上がるように、予測音声含有率を下記のように調整してもよい。

【数１３】

$$W_{(z)} = W_{(z-1)} + \gamma(Y_{(z-1)} - Y_{(z)}) \qquad \cdots (13)$$

Ｗ（ｚ）は、セクションＺにおける、調整後の予測音声含有率である。以下、調整音声含有率と呼ぶ。Ｗ（ｚ－１）は、セクションＺの１つ前のセクションにおける調整音声含有率である。Ｙ（ｚ）は、セクションＺにおける予測音声含有率である。Ｙ（ｚ－１）は、１つ前のセクションにおける予測音声含有率である。γは、加算する際の係数である。

【０１０８】

この調整音声含有率を用いて速度比条件を設定することで、予測音声含有率が１つ前のセクションにおける予測音声含有率よりも上がれば、音声区間の平均速度比が下がる。また予測音声含有率が１つ前のセクションにおける予測音声含有率と同じ場合、音声区間の平均速度比は変化しない。

【０１０９】

このような調整音声含有率を利用することで、音声含有率の高いセクションではより遅くすることが可能となり、情報量に応じた速度条件の設定が可能となる。なお、このような調整を行うことで、セクション圧伸比を達成できずに誤差が生じる恐れがあるが、式（１２）に示すように、次以降のセクションにおいて誤差を解消することができる。これにより、目標圧伸比は達成することができる。例えば、図１９に示したように、予測音声含有率は一定ではなく、高いところもあれば低いところもある。そのため、予測音声含有率が高いセクションで再生速度が遅くなり、目標圧伸比と差が広がったとしても、次以降の予測音声含有率が低いセクションでこの差を解消することができる。

【０１１０】

音声区間長予測部１９は、音声非音声判別部１１の過去の判別結果から、音声区間長を算出し、音声区間長の予測を行う。音声区間長の予測値は、一つ一つの音声区間長の予測値ではなく、文を話す際の平均的な音声区間長を代表値として予測する。一つ一つの音声区間長は、話者交替や会話内容などの様々な要因によって、予測することは難しい。そこで、平均的な音声区間長を予測値として利用して、コンテンツに適した音声区間の速度比制御を行う。図９及び図１０に示したように、音声区間長の分布はジャンル毎に大きく異なる。図９は、ニュース番組での各音声区間長の頻度を示したものであるが、５００ｍｓをピークとし、４０００ｍｓ辺りまでゆるやかに頻度が減っている。一方、図１０は、野球番組での各音声区間長の頻度を示したものであるが、同じく５００ｍｓをピークとしているものの、急激に頻度が減少していく様子がみられる。この減少の仕方の違いにより、これらの番組を視聴した際の印象としては、ニュース番組では長めの音声区間が続き、野球番組では短い音声区間が続くように聞こえる。そのため、音声区間長を予測せずに、固定長により音声区間の速度比制御を行うと、音声区間の速度比が必要以上に遅くなりすぎたり、速すぎたりする恐れがある。これは、音声非音声判別部１１が逐次的に音声か非音声かを判別しているために、音声区間の終端時刻が把握できないからである。また、音声区間長の違いは、話者や会話内容の違いによるものであり、コンテンツ毎に異なるものとなる。このような理由から、音声区間長の予測が必要となる。

【０１１１】

そこで、処理を開始してからｎ（ｎは自然数）番目の音声区間がもつ音声区間長の予測値Ｌ（ｎ）は、一つ前のｎ－１番目の音声区間がもつ音声区間長の予測値Ｌ（ｎ－１）と実測値Ｍ（ｎ－１）とから、式（１４）で表現される。

【数１４】

$$L_{(n)} = \frac{n-1}{n} L_{(n-1)} + \frac{1}{n} M_{(n-1)} \qquad \cdots (14)$$

なお、音声区間には、「はい」や「うん」など相槌や、「えー」などのフィラーなどが含まれる。これらは言語的な理解が容易なため、速度比に関わらず聞き取り易い。そこで、式（１４）の音声区間長の予測値Ｌ（ｎ）の算出には、所定の閾値以上の音声区間長をもつ音声区間を利用するとした。ここでは、所定の閾値の一例として、１０００ｍｓを採用する。

【０１１２】

式（１４）に基づいて予測した音声区間長を図２０に示す。図２０において、縦軸は音声区間長であり、横軸は経過時間を示している。図２０では、Ｘ軸において音声区間長の始端時刻の位置に、その音声区間長を示している。また図２０では、音声区間長の実測値Ｍ（ｎ）、直前の音声区間長の実測値Ｍ（ｎ－１）、及び予測音声区間長Ｌ（ｎ）と図示している。このように、相槌やフィラーなどを除いた音声区間長から予測音声区間長を算出することで、速い速度比では聞き取りにくくなる音声区間長が長い音声区間に適した速度比を設定することができる。また、コンテンツに応じて音声区間長が逐次的に算出されることで、音声区間長の分布の違いに応じた速度比の設定が可能となる。速度比決定部１６は、逐次的な速度比条件と予測音声区間長とに基づき、音声及び非音声区間の変換速度

比を決定する。

【０１１３】

以下、図２１を参照して、第２の実施形態に係る音声再生装置の処理について説明する。図２１は、第２の実施形態に係る音声再生装置の処理の流れを示すフローチャートである。

【０１１４】

まず、入力装置（図示なし）においてユーザによるコンテンツを再生する指示を受け付けたか否かが判断される（ステップＳ２０１）。ユーザがコンテンツを再生する指示をしたとき、音声非音声判別部１１にオーディオ信号が入力される。音声非音声判別部１１は、入力されたコンテンツのオーディオ信号について音声区間と非音声区間とをフレーム毎に判別する（ステップＳ２０２）。

【０１１５】

ステップＳ２０２の次に、音声含有率予測部１８は、音声非音声判別部１１から出力されたフレーム毎の判別結果から、現時点より過去数分の音声含有率を算出し、算出した音声含有率を用いて、現時点より１つ進んだセクションにおける音声含有率を予測する（ステップＳ２０３）。

【０１１６】

ステップＳ２０３の次に、速度比条件設定部１４は、音声含有率予測部１８で算出された予測音声含有率と、圧伸比算出部２０で算出されたセクション圧伸比を入力とし、セクション毎に速度比条件を設定する（ステップＳ２０４）。また、音声区間長予測部１９は、音声非音声判別部１１の過去の判別結果から、音声区間長を算出し、音声区間長の予測を行う（ステップＳ２０５）。

【０１１７】

ステップＳ２０５の次に、速度比決定部１６は、音声非音声判別部１１から出力される音声区間の始終端時刻を参照して、所定の単位時間毎に音声区間であるか否かを判断する（ステップＳ２０６）。音声区間と判断した場合、速度比決定部１６は、音声区間における経過割合を算出する（ステップＳ２０７）。音声区間の経過割合とは、音声区間の始端を０、終端を１として、始端からの経過時間を音声区間長で除算したものである。本実施形態では、音声非音声判別が逐次的に行われているため、音声区間の始端時刻は把握できるが、音声区間の終端時刻は現時点では分からない。そこで、音声区間長予測部１９で予測された予測音声区間長を音声区間長として用いる。これにより、速度比決定部１６は、音声区間における経過割合を算出することができる。なお、音声区間長として実際の値ではなく、予測音声区間長を用いるので、実際の音声区間の経過割合とは必ずしも一致しない。従って、音声区間の経過割合が１以下であっても音声区間の終端時刻となる可能性がある。そこで、速度比決定部１６は、音声区間の経過割合が１を越えていないかどうかを判断する（ステップＳ２０８）。１を超えていない場合、速度比決定部１６は、音声区間の経過割合から音声区間の変換速度比を決定する（ステップＳ２０９）。ステップＳ２０９の処理は、第１の実施形態の速度比決定部１６で説明した処理と同様であるので、説明を省略する。ステップＳ２０８において経過割合が１を超えた場合、処理はステップＳ２１０へ進み、速度比条件である終端速度比を変換速度比に算出する。この場合、音声区間長予測部１９で予測された音声区間長を超過した状態であるため、音声区間の終端速度比を変換速度比として算出する必要がある。

【０１１８】

ステップＳ２０６において非音声区間と判断した場合、速度比決定部１６は、非音声区間の変換速度比を決定する（ステップＳ２１１）。ステップＳ２０９、Ｓ２１０、及びＳ２１１の次に、速度比決定部１６は、速度変換対象となるコンテンツの終端時刻まで変換速度比を決定したか否かを判断する（ステップＳ２１２）。終端時刻ではないとき、処理はステップＳ２０２へ戻る。このように、速度変換対象となるコンテンツの終端時刻までの変換速度比が算出されるまで、速度比決定部１６においてステップＳ２０２～Ｓ２１２までの処理がセクション単位で繰り返される。ステップＳ２１２においてコンテンツの終

端時刻まで変換速度比が算出されたと判断された場合、入力装置（図示なし）が本装置の処理を終了するか否かの指示を受け付ける（ステップＳ２１３）。ユーザが他のコンテンツについて速度変換処理を行う場合（ステップＳ２１３でＮｏ）、処理はステップＳ２０２へ戻る。

【０１１９】

以上のように、本実施形態に係る音声再生装置によれば、リアルタイムで処理を行いながら速度変換を行うことができる。また第１の実施形態に比べ、音声含有率や音声区間長に予測値を用いている。このため、実測値との誤差が生じるが、この誤差は圧伸比設定部２８で設定されるセクション圧伸比によって解消される。これにより、本実施形態に係る音声再生装置によれば、リアルタイムで処理を行いながら、目標圧伸比を達成しつつ、区間削除や音声区間の極端な高速化をせずに、速度変換を行うことができる。

【０１２０】

（第３の実施形態）

図２２を参照して、本発明の第３の実施形態に係る音声再生装置について説明する。図２２は、第３の実施形態に係る音声再生装置の構成例を示すブロック図である。図２２において、本音声再生装置は、音声非音声判別部１１、音声含有率予測部１８、速度比条件設定部１４、音声区間長予測部１９、速度比決定部１６、速度変換部１７、圧伸比算出部２０、及び統計量算出部２１で構成される。本実施形態は、第２の実施形態に係る音声再生装置に対し、統計量算出部２１を新たに備え、速度比条件設定部１４の処理が異なる。以下、統計量算出部２１と、速度比条件設定部１４の処理を中心に説明する。

【０１２１】

統計量算出部２１は、音声区間の上限速度比を修正するための統計量を算出している。例えば、コンテンツの始端から現時点までの音声含有率を利用する。このような音声含有率を以下、長期音声含有率と称す。コンテンツ毎の長期音声含有率の時間変化を図２３に示す。図２３において、縦軸は長期音声含有率を示し、横軸は始端（０分）からの経過時間を示している。また、コンテンツ毎の予測音声含有率を図２４に示す。予測音声含有率は、第２の実施形態において説明した予測音声含有率と同じである。ここでは、算出間隔を１分としている。予測音声含有率のグラフは、音声区間が密集している部分や疎の部分が反映され、山谷がはっきりしたグラフとなっている。長期音声含有率のグラフは、始端付近で多少変動があるものの、概ね平坦であり、第１の実施形態で用いたコンテンツ全体に対する音声含有率に近いグラフとなる。

【０１２２】

そこで、この長期音声含有率を用いて音声区間の上限速度比の修正を行うことを考える。音声区間の上限速度比の修正を逐次行っていくことで、予測音声含有率が局所的に高くなった場合でも、音声区間の速度比が上がりすぎることを防ぐことができる。

【０１２３】

速度比条件設定部１４は、音声含有率予測部１８で算出された予測音声含有率、統計量算出部２１で算出された長期音声含有率、及び圧伸比算出部２０で算出されたセクション圧伸比を入力とする。上述した図２１のステップＳ２０４において、速度比条件設定部１４は図２５に示す処理を行う。図２５は、第３の実施形態に係る速度比条件設定部１４の処理を示すフローチャートである。

【０１２４】

図２５において、速度比条件設定部１４は、入力される音声含有率が予測音声含有率であるか否かを判断する（ステップＳ３０１）。予測音声含有率が入力された場合、処理はステップＳ３０２へ進み、速度比条件設定部１４は、予測音声含有率を用いて音声及び非音声の平均速度比を算出する。また速度比条件設定部１４は、算出した音声区間の平均速度比から終端速度比を算出する（ステップＳ３０３）。なお、予測音声含有率に基づく終端速度比をＶｅｎｄ１とする。ステップＳ３０２及びＳ３０３の処理は、上述した第１の実施形態と同様の処理である。

【０１２５】

一方、ステップＳ３０１において予測音声含有率が入力されない場合、つまり長期音声含有率が入力された場合、速度比条件設定部１４は、長期音声含有率を用いて音声及び非音声の平均速度比を算出する（ステップＳ３０４）。また速度比条件設定部１４は、算出した音声区間の平均速度比から終端速度比を算出する（ステップＳ３０５）。なお、長期音声含有率に基づく終端速度比をＶｅｎｄ２とする。ステップＳ３０４及びＳ３０５の処理は、上述した第１の実施形態と同様の処理である。

【０１２６】

ステップＳ３０５の次に、速度比条件設定部１４は、長期音声含有率に基づいて算出した音声区間の終端速度比Ｖｅｎｄ２を上限速度比として設定する（ステップＳ３０６）。ステップＳ３０３及びＳ３０６の次に、速度比条件設定部１４は、予測音声含有率に基づく終端速度比Ｖｅｎｄ１と、長期音声含有率に基づく上限速度比Ｖｅｎｄ２とを比較する（ステップＳ３０７）。終端速度比Ｖｅｎｄ１が上限速度比Ｖｅｎｄ２を超える場合（Ｖｅｎｄ１＞Ｖｅｎｄ２）、速度比条件設定部１４は終端速度比Ｖｅｎｄ１を上限速度比Ｖｅｎｄ２に修正する（ステップＳ３０８）。またこの修正に併せて、速度比条件設定部１４は音声区間の平均速度比も長期音声含有率によって算出された値に修正する。つまり、音声及び非音声区間の平均速度比、音声区間の終端速度比の３つの速度比を表す速度比条件のうち、非音声区間の平均速度比についてのみ予測音声含有率で算出された値を用いる。それ以外の音声区間の平均速度比及び終端速度比は、長期音声含有率によって求められた値を用いる。

【０１２７】

このように、音声区間の平均速度比及び終端速度比の修正を逐次行っていくことで、予測音声含有率が局所的に高くなって音声区間の平均速度比が高めに設定され得る場合でも、聞き易い速度比での再生を行うことができる。

【０１２８】

音声区間長予測部１９は、音声非音声判別部１１の過去の判別結果から、予測音声区間長を算出する。本実施形態では、予測音声区間長として音声区間長の最大値を利用する。これは、聞き取り易さ重視の観点から、どのような音声区間であっても漏らさずに終端まで速度比制御を行えるようにするためである。

【０１２９】

図２６に示すように、音声区間長の分布は経過時間によって大きく異なる。このため、音声区間長の予測が必要となる。図２６は、音声区間長の実測値と、直前の音声区間長の実測値と、予測音声区間長の分布を示した図である。図２６において、縦軸は音声区間長を示し、横軸はコンテンツの始端からの経過時間を示している。また音声区間長は音声区間の始端時刻に表示している。ここで、ｎ番目の音声区間長の実測値をＭ（ｎ）とする。予測音声区間長Ｌｍ（ｎ）は式（１５）～式（１７）のように表現される。

【数１５】

$$Lm_{(1)} = \max \quad \cdots (15)$$

【数１６】

$$Lm_{(n)} = Lm_{(n-1)} + \beta \quad , \quad M_{(n-1)} < Lm_{(n-1)} \quad \cdots (16)$$

【数１７】

$$Lm_{(n)} = M_{(n-1)} \quad , \quad M_{(n-1)} \geqq Lm_{(n-1)} \quad \cdots (17)$$

式（１５）において、ｍａｘはコンテンツに含まれる音声区間長のうち最大の音声区間長を複数のコンテンツについて平均した値である。事前にジャンル情報が得られる場合は、ジャンル情報毎に上記平均値を算出し、テーブルを用意しておく。

【０１３０】

βは、予測音声区間長Ｌｍが次の音声区間までの経過時間とともに減少するように設定された値である。ｎ－１番目の音声区間の始端時刻を０とし、ｎ番目の音声区間の始端時刻をｔとすると、式（１８）のように表せる。ｋは正の値をとるものとする。

【数１８】

$$\beta = -kt \quad \cdots(18)$$

なお、βは指数関数でもよく、経過時間ｔの減少関数であればよい。

【０１３１】

式（１５）～式（１８）により予測された予測音声区間長は、図２６に示すようになる。図２６に示すように、予測音声区間長が音声区間長の実測値よりも長いものが多い。この値を速度比算出時に用いることで、音声区間の終端時刻では終端速度比で変換される割合が低下し、音声区間の平均速度比が更に下がる効果を有する。その結果、聞き取り易い再生を提供することができる。

【０１３２】

（第４の実施形態）

図２７を参照して、本発明の第４の実施形態に係る音声再生装置について説明する。図２７は、第４の実施形態に係る音声再生装置の構成例を示すブロック図である。図２７において、本音声再生装置は、音声非音声判別部１１、一時蓄積部２２、音声含有率算出部１３、速度比条件設定部１４、音声区間長算出部１５、速度比決定部１６、速度変換部１７、及び圧伸比算出部２０で構成される。本実施形態は、第１の実施形態に係る音声再生装置に対し、蓄積部１２よりも蓄積量が少ない一時蓄積部２２と、第２の実施形態で説明した圧伸比算出部２０を備える点で異なる。以下、異なる点を中心に説明する。

【０１３３】

一時蓄積部２２は、ハードディスク、ＤＶＤ、又はメモリ媒体（例えばＳＤカード）などの読み書き可能な記録媒体で構成される。一時蓄積部２２には、音声非音声判別部１１に入力されるのと同じオーディオ信号がセクション１個分もしくは数個分蓄積される。そして、一時蓄積部２２において蓄積されたオーディオ信号が速度変換処理された後、その速度変換処理されたセクションのオーディオ信号は消去され、新しいセクションのオーディオ信号が蓄積される。ここで、本実施形態に係るセクションとは、所定間隔で区切られた区間だけではなく、所定のイベントで区切られた区間でもよい。例えば、イベントをＣＭとすると、ＣＭ区間と、ＣＭとＣＭに挟まれた番組区間の２種類のセクションができる。イベントが音楽であれば、音楽区間と、音楽と音楽に挟まれた区間の２種類のセクションができる。また、セクションは、ユーザによって指示された区間であってもよい。

【０１３４】

なお、第１の実施形態と同様、１つのセクションを構成するオーディオ信号及びビデオ信号が一時蓄積部２２に蓄積されるとき、音声非音声判別部１１において判別処理が行われ、当該セクションの判別結果や音声区間の始終端時刻も一時蓄積部２２に蓄積される。また、一時蓄積部２２には、オーディオ信号やビデオ信号と、判別結果及び音声区間の始終端時刻とが対応付けされて蓄積される。なお、オーディオ信号及びビデオ信号のフォーマットは、どのようなフォーマットであってもかまわない。また、本実施形態に係る音声再生装置が、上述した蓄積部１２をさらに備えていてもよい。この場合、蓄積部１２においてコンテンツ単位で蓄積されたオーディオ信号や音声非音声判別結果が、セクション単位で読み出され、一時蓄積部２２に蓄積されるようにする。

【０１３５】

このような一時蓄積部２２を設けることで、蓄積されたセクションで実際の音声区間長（実測値）を算出することができる。これにより、実際の音声区間にあわせた速度制御が可能になる。また蓄積されたセクションの実際の音声含有率を求めることができるため、第２の実施形態で説明した予測音声含有率を用いる場合に比べて、局所的な変動が少なく、コンテンツ全体の音声含有率と近い値となる。

【０１３６】

音声含有率算出部１３は、一時蓄積部２２で蓄積された判別結果や音声区間の始終端時刻から、セクションの音声含有率を算出する。このセクション内に含まれる音声区間長の和を求め、セクション全体の時間長（以下、セクション長と称す）で除算したものが本実施形態の音声含有率となる。

【０１３７】

以下、図２８を参照して、第４の実施形態に係る音声再生装置の処理について説明する。図２８は、第４の実施形態に係る音声再生装置の処理の流れを示すフローチャートである。

【０１３８】

まず、入力装置（図示なし）において、ユーザが所望のコンテンツを再生する指示をしたか否かが判断される（ステップＳ４０１）。ユーザの指示があった場合、コンテンツのオーディオ信号及びビデオ信号がセクション分だけ一時蓄積部２２に蓄積され、音声非音声判別部１１は、セクション内のオーディオ信号について音声区間と非音声区間とを判別する（ステップＳ４０２）。なお、ステップＳ４０２において判別された判別結果と音声区間の始終端時刻についても、一時蓄積部２２に蓄積される。

【０１３９】

ステップＳ４０２の次に、音声含有率算出部１３は、セクションの音声含有率を算出する（ステップＳ４０３）。速度比条件設定部１４は、ステップＳ４０３で算出された音声含有率、圧伸比算出部２０で算出されたセクション圧伸比に基づいて、音声区間の平均速度比、非音声区間の平均速度比、及び音声区間の終端速度比を算出する（ステップＳ４０４及びＳ４０５）。この処理は、第１の実施形態と同様の処理である。次に、ステップＳ４０６において速度比条件設定部１４は、ステップＳ４０５で算出した終端速度比Ｖｅｎｄ１と、ユーザによって指定された又は予め装置に設定された終端速度比の上限速度比Ｖｅｎｄ２とを比較する。ステップＳ４０６において終端速度比Ｖｅｎｄ１が上限速度比Ｖｅｎｄ２よりも大きいと判断された場合、速度比条件設定部１４は、終端速度比Ｖｅｎｄ１を上限速度比Ｖｅｎｄ２に修正する（ステップＳ４０７）。ステップＳ４０６において終端速度比Ｖｅｎｄ１が上限速度比Ｖｅｎｄ２よりも小さいと判断された場合、処理はステップＳ４０８へ進む。

【０１４０】

ここで、セクション音声含有率は、コンテンツの一部を構成するセクション内での値である。したがって、音声区間が局所的に集中するセクションなどが存在すれば、局所的にセクション音声含有率の値が大きくなる場合がある。以上のステップＳ４０６及びＳ４０７の処理を行うことで、セクション音声含有率の値が大きくなり、音声区間の終端速度比が大きくなり過ぎることを防ぐことができる。なお、第３の実施形態で説明した統計量算出部２１で長期音声含有率を算出し、上限速度比を修正するようにしてもよい。

【０１４１】

ステップＳ４０７の次に、音声区間長算出部１５は、音声区間の始終端時刻を入力とし、音声区間長を算出する（ステップＳ４０８）。速度比決定部１６は、一時蓄積部２２に蓄積された音声区間の始終端時刻を参照して、セクションの始端から順に所定の単位時間毎に音声区間であるか否かを判断する（ステップＳ４０９）。音声区間と判断した場合、速度比決定部１６は、音声区間の始終端時刻と、ステップＳ４０８で算出された音声区間長とに基づき、音声区間における経過割合を算出する（ステップＳ４１０）。

【０１４２】

ステップＳ４１０の次に、速度比決定部１６は、音声区間の経過割合から、音声区間の

変換速度比を決定する（ステップＳ４１１）。ステップＳ４１１の処理は、第１の実施形態と同様である。ステップＳ４０９において非音声区間と判断した場合、速度比決定部１６は、非音声区間の始端から終端まで、速度比条件設定部１４で設定された非音声区間の平均速度比を変換速度比として決定する（ステップＳ４１２）。

【０１４３】

ステップＳ４１１及びＳ４１２の次に、速度比決定部１６は、セクションの終端まで変換速度比を算出したか否かを判断する（ステップＳ４１３）。終端ではないとき、処理はステップＳ４０９へ戻る。このように、セクションの終端までの変換速度比が算出されるまで、速度比決定部１６においてステップＳ４０９～Ｓ４１３までの処理が繰り返される。ステップＳ４１３においてセクションの終端まで変換速度比が算出されたと判断された場合、速度変換部１７において変換速度比に従ってオーディオ信号の速度変換が行われ、速度変換後のオーディオ信号の再生が開始される（ステップＳ４１４）。速度比決定部１６は、速度変換対象となるコンテンツの終端時刻まで再生されたか否かを判断する（ステップＳ４１５）。終端時刻ではないとき、次のセクション分のオーディオ信号が一時蓄積部２２に蓄積され、処理はステップＳ４０２へ戻る。ステップＳ４１５においてコンテンツの終端時刻まで再生されたと判断された場合、入力装置（図示なし）が本装置の処理を終了するか否かの指示を受け付ける（ステップＳ４１６）。ユーザが他のコンテンツについて速度変換処理を行う場合（ステップＳ４１６でＮｏ）、処理はステップＳ４０１へ戻る。

【０１４４】

以上のように、本実施形態に係る音声再生装置によれば、セクション単位で速度変換を行うことができる。ここでユーザが、例えば放送中の番組を最初から録画していたが、その番組放送の途中から視聴可能になったとする。このときユーザは、その番組の冒頭を見逃したのでその冒頭を速度変換して視聴しようとするとき、ユーザは速度変換処理の開始時点を冒頭の時点に指定する。なお、速度変換処理の終了時点は、最新の録画がなされた時点である。つまり、冒頭から最新の録画がなされた時点までの区間が１つのセクションとなる。これにより、ユーザは、冒頭から最新の録画がなされた時点まで速度変換した視聴をすることができ、その後においては通常再生によって視聴を続けることができる。このように、本実施形態に係る音声再生装置によれば、セクション単位で速度変換処理を行うので、コンテンツの録画中であっても、全体の録画終了を待たずに速度変換処理を行うことができる。

【０１４５】

また本実施形態に係る音声再生装置によれば、一時蓄積部２２を備えることにより、音声含有率の実測値を算出することができ、音声含有率として予測値を用いる第２及び第３の実施形態に比べて、より最適な速度比で速度変換を行うことができる。また、一時蓄積部２２を備えることにより、音声区間長の実測値を算出することができる。音声区間長に実測値を用いる限り、音声区間長が分からないことによる終端速度比の上がり過ぎは生じず、音声区間長として予測値を用いる第２及び第３の実施形態に比べて、より最適な速度比で速度変換を行うことができる。

【０１４６】

（第５の実施形態）

図２９を参照して、本発明の第５の実施形態に係る音声再生装置について説明する。図２９は、第５の実施形態に係る音声再生装置の構成例を示すブロック図である。図２９において、本音声再生装置は、音判別部２３、蓄積部１２、音声含有率算出部１３、速度比条件設定部１４、音声区間長算出部１５、速度比決定部１６、速度変換部１７、及び特定イベント含有率算出部２４で構成される。

【０１４７】

なお、上述した第１の実施形態では、音声区間及び非音声区間の速度比を算出したが、本実施形態では、コンテンツに含まれる特定イベント区間についてさらに個別の速度比を算出することが可能な音声再生装置について説明する。また本実施形態に係る音声再生装

置は、第１の実施形態に係る音声再生装置に対し、音声非音声判別部１１の代わりに音判別部２３を備える点と、特定イベント含有率算出部２４をさらに備える点で大きく異なる。

【０１４８】

音判別部２３は、オーディオ信号を入力として、特定イベント音を含む特定イベント区間、当該特定イベント区間以外の音声区間及び非音声区間を判別する。特定イベント音とは、個別の音源からの音であってもよいし、複数の音源からの音を一まとめにしたものであってもよい。個別の音源からの音としては、例えば、話者Ａからの音声、楽器Ｂからの音、機器Ｃからの特定音などが挙げられる。複数の音源からの音を一まとめにしたものとしては、例えば、複数の話者からの音声を一まとめにしたものや、音楽、雑音などが挙げられる。また、特定イベント音は１つとは限らず、複数であってもよい。特定イベント音が複数ある場合、音判別部２３は、オーディオ信号を入力として、複数の特定イベント区間、当該各特定イベント区間以外の音声区間及び非音声区間を判別することになる。また、特定イベント音が話者Ａや話者Ｂなどの音声である場合、音判別部２３が判別する音声区間は特定イベント区間以外の音声区間を意味することになる。以下では、特定イベント音を音楽と仮定して説明する。

【０１４９】

音判別部２３は、オーディオ信号を入力として、特定イベント区間である音楽区間、当該音楽区間以外の音声区間及び非音声区間を判別する。これらの区間を判別する方法としては、例えば「ＭＰＥＧ符号化データからのオーディオインデキシング」＜中島康之，陸洋，　菅野勝，　柳原広昌，　米山暁夫（ＫＤＤＩ研究所）２０００、信学論　Ｄ－II，　Ｖｏｌ．Ｊ８３－Ｄ－II，　Ｎｏ．５，　ｐｐ．１３６１－１３７１＞に記載された公知の方法がある。この方法では、まずオーディオ信号を有音部と無音部に分類する。そして有音部についてさらに、ベイズ推定を用いて音声・音楽・歓声の３つのカテゴリに分類する。このような方法で、音判別部２３は、音楽区間、当該音楽区間以外の音声区間及び非音声区間を判別する。なお、上記歓声は、音楽区間以外の非音声区間に含まれるとする。音判別部２３で判別された判別結果や、音声区間の始終端時刻、音楽区間の始終端時刻は、蓄積部１２に蓄積される。

【０１５０】

特定イベント含有率算出部２４は、蓄積部１２に蓄積された特定イベント区間の始終端時刻から特定イベントの含有率を算出する。特定イベント含有率は、コンテンツのオーディオ信号に含まれる特定イベント区間（ここでは音楽区間）の比率を示したものである。以下の説明では、特定イベント含有率を音楽含有率と言い換えて説明する。音楽含有率は、具体的には、所定時間のオーディオ信号に含まれる音楽区間長の和を当該所定時間で除算したものである。ここでは、コンテンツ全体に含まれる音楽区間長の和をコンテンツ長で除算したものとする。

【０１５１】

速度比条件設定部１４は、まず目標圧伸比と特定イベント含有率算出部２４で算出された音楽含有率から、音楽区間の平均速度比（以下、音楽速度比と称す）を算出する。なお、目標圧伸比は、ユーザによって設定されたものでもよいし、予め装置に設定されたものでもよい。具体的には、速度比条件設定部１４は、目標圧伸比に応じて異なる音楽含有率と音楽速度比との対応を示したテーブルや対応関数に基づいて、音楽速度比を算出する。このテーブルや対応関数は、予め用意されているとする。図３０は、音楽含有率と音楽速度比との対応を示したテーブルの例を示す図である。図３０に示すテーブルは、目標圧伸比が０．５のときの対応関係を示したものである。

【０１５２】

速度比条件設定部１４は、算出した音楽速度比に基づいて、音楽区間以外の音声区間及び非音声区間の圧伸比を算出する。ここで音楽含有率をＳｍ、音楽速度比をＦ、目標圧伸比をＥ、音楽区間以外の音声区間及び非音声区間の平均速度比をＧとすると、平均速度比Ｇは式（１９）となる。

【数１９】

$$G=\frac{F-FSm}{FE-Sm} \quad \cdots(19)$$

例えば目標圧伸比が０．５、音楽含有率が１０％の場合、図３０により、音楽速度比は１倍速となる。したがってこの場合、式（１９）にＥ＝０．５、Ｓｍ＝０．１、Ｆ＝１を代入すると、Ｇ＝２．２５となる。

【０１５３】

圧伸比は速度比の逆数で表せる。平均速度比Ｇが２．２５であるため、音楽区間以外の音声区間及び非音声区間の圧伸比はその逆数０．４４となる。そこで、この圧伸比（０．４４）を音声区間及び非音声区間についての目標圧伸比とすれば、第１の実施形態と同様の方法で、音声区間の平均速度比、非音声区間の平均速度比、音声区間の終端速度比を算出することができる。なお、速度比条件設定部１４が用いる速度比算出分布は、音楽含有率に応じて設定されるようにしてもよい。

【０１５４】

以下、図３１を参照して、第５の実施形態に係る音声再生装置の処理について説明する。図３１は、第５の実施形態に係る音声再生装置の処理の流れを示すフローチャートである。

【０１５５】

まず、ユーザが入力装置（図示なし）においてコンテンツを録画する指示をしたとき、当該コンテンツのオーディオ信号及びビデオ信号が蓄積部１２に蓄積される。このとき、音判別部２３は、音楽区間、当該音楽区間以外の音声区間及び非音声区間を判別する（ステップＳ５０１）。なお、ステップＳ５０１において判別された判別結果、音声区間の始終端時刻、及び音楽区間の始終端時刻についても、蓄積部１２に蓄積される。

【０１５６】

ステップＳ５０１の次に、入力装置において、ユーザが所望のコンテンツを再生する指示をしたか否かが判断される（ステップＳ５０２）。ユーザの指示があった場合（ステップＳ５０２でＹｅｓ）、音声含有率算出部１３は、指示されたコンテンツの音声含有率を算出する（ステップＳ５０３）。また、特定イベント含有率算出部２４は、指示されたコンテンツの音楽含有率を算出する（ステップＳ５０４）。

【０１５７】

ステップＳ５０４の次に、速度比条件設定部１４は、目標圧伸比と特定イベント含有率算出部２４で算出された音楽含有率から、音楽速度比を算出する（ステップＳ５０５）。速度比条件設定部１４は、算出した音楽速度比に基づいて、式（１９）を用いて音楽区間以外の音声区間及び非音声区間の圧伸比を算出する。そして、算出した圧伸比を用いて、音声区間の平均速度比、非音声区間の平均速度比、及び音声区間の終端速度比を速度比条件として設定する（ステップＳ５０６）。音声区間長算出部１５は、音声区間の始終端時刻を入力とし、音声区間長を算出する（ステップＳ５０７）。

【０１５８】

ステップＳ５０７の次に、速度比決定部１６は、蓄積部１２に蓄積された音楽区間の始終端時刻を参照して、コンテンツの始端から順に所定の単位時間毎に音楽区間であるか否かを判断する（ステップＳ５０８）。音楽区間と判断した場合、速度比決定部１６は、ステップＳ５０５で算出した音楽速度比を変換速度比として決定する（ステップＳ５０９）。つまり、音楽区間の始端から終端までの変換速度比は、音楽速度比で一定となる。

【０１５９】

ステップＳ５０８において音楽区間でないと判断した場合、速度比決定部１６は、音声区間の始終端時刻を参照して、音声区間であるか否かを判断する（ステップＳ５１０）。音声区間と判断した場合、速度比決定部１６は、ステップＳ５０６で設定された音声区間の平均速度比と、音声区間の始終端時刻と、ステップＳ５０７で算出された音声区間長と

に基づき、音声区間における経過割合を算出する（ステップＳ５１１）。速度比決定部１６は、音声区間の経過割合から、音声区間の変換速度比を決定する（ステップＳ５１２）。音声区間でないと判断した場合、速度比決定部１６は、ステップＳ５０６で設定された非音声区間の平均速度比を変換速度比として決定する（ステップＳ５１３）。つまり、非音声区間の始端から終端までの変換速度比は、当該非音声区間の平均速度比で一定となる。なお、ステップＳ５１１～Ｓ５１３の処理は、第１の実施形態と同様である。

【０１６０】

ステップＳ５０９、Ｓ５１２、及びＳ５１３の次に、速度比決定部１６は、コンテンツの終端まで変換速度比を算出したか否かを判断する（ステップＳ５１４）。終端ではないとき、処理はステップＳ５０８へ戻る。このように、コンテンツの終端までの変換速度比が算出されるまで、速度比決定部１６においてステップＳ５０８～Ｓ５１４までの処理が繰り返される。ステップＳ５１４においてコンテンツの終端まで変換速度比が算出されたと判断された場合、速度変換部１７において変換速度比に従ってオーディオ信号の速度変換が行われ、速度変換後のオーディオ信号の再生が開始される（ステップＳ５１５）。入力装置（図示なし）が本装置の処理を終了するか否かの指示を受け付ける（ステップＳ５１６）。ユーザが他のコンテンツについて速度変換処理を行う場合（ステップＳ５１６でＮｏ）、処理はステップＳ５０２へ戻る。

【０１６１】

以上のように、本実施形態に係る音声再生装置によれば、特定イベント含有率を算出して特定イベント区間の速度比を設定することで、音楽番組などを視聴するに際し、特定イベント区間である音楽区間をそれ以外の音声区間及び非音声区間よりも遅い速度で再生することができる。これにより、速度変換処理において、音楽を重視した再生を行うことができる。また、特定イベント音を音楽ではなく、コンテンツ中に登場するある話者Ａの音声とした場合、話者Ａの音声に重点がおかれ、特定イベント区間以外の音声区間よりも遅い速度で話者Ａの音声が速度変換処理される。例えば、何度も視聴しているコンテンツに対して、話者Ａの発言内容を確認したいときなどに、話者Ａの発言内容だけ遅い速度で再生を行うことは有用である。また、セキュリティカメラのように長時間記録し続けている場合、雑音部分を特定イベント音として識別し、その部分を高速再生することで、冗長なシーンを見る時間を減らすような使い方も可能となる。このように、ある特定イベント音に対して、個別の速度を設定することにより、その部分の重視を促すように遅い速度で再生したり、冗長な部分を低減するために速い速度で再生を行ったり、用途に応じた速度設定が可能になる。

【０１６２】

なお、上述した第１～第５の実施形態に係る音声再生装置は、一般的なコンピュータシステム５０に音声再生プログラムを実行させることによって実現されてもよい。図３２は、音声再生装置がコンピュータシステム５０によって実現される構成例を示すブロック図である。

【０１６３】

図３２において、コンピュータシステム５０は、ＣＰＵ５１、メモリ５２、ハードディスク５３、ディスクドライブ装置５４、モニタ５５、スピーカ５６、及び入力装置５７で構成される。ＣＰＵ５１は、音声再生プログラムを実行させることによって、上述した蓄積部１２及び一時蓄積部２２以外の第１～第５の実施形態に係る音声再生装置の各構成部と同一の機能を実現する。メモリ５２やハードディスク５３は、音声再生プログラムを実行させることによって、蓄積部１２及び一時蓄積部２２と同一の機能を実現する。

【０１６４】

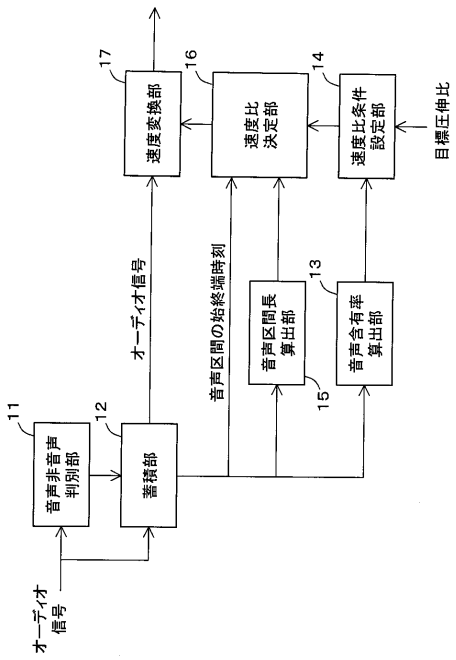
ディスクドライブ装置５４は、コンピュータシステム５０を音声再生装置として機能させるための音声再生プログラムが記憶された記録媒体５８から、当該音声再生プログラムを読み出す。音声再生プログラムが任意のコンピュータシステム５０にインストールされることにより、コンピュータシステム５０を上述した音声再生装置として機能させることができる。

【０１６５】
なお、記録媒体５８は、例えばフレキシブルディスクや光ディスクなどのディスクドライブ装置５４によって読み取り可能な形式の記録媒体である。また音声再生プログラムは、コンピュータシステム５０に予めインストールされていてもかまわない。また音声再生プログラムは、インターネットなどの電気通信回線によって提供されてもよい。また音声再生処理は、全部または一部をハードウェアによって処理される形態であってもよい。
【０１６６】
モニタ５５は、ディスクドライブ装置５４を介して読み込んだ記録媒体５８に記録されたビデオ信号や、ハードディスク５３に記録されたビデオ信号などを表示する。スピーカ５６は、ディスクドライブ装置５４を介して読み込んだ記録媒体５８に記録されたオーディオ信号、ハードディスク５３に記録されたオーディオ信号、速度変換処理後のオーディオ信号を音に変換して再生する。入力装置５７は、例えばキーボードやマウスなどで構成され、目標圧伸比の入力などを受け付ける。
【０１６７】
このように、上述した第１～第５の実施形態に係る音声再生装置は、一般的なコンピュータシステム５０に音声再生プログラムを実行させることによって実現される。
【０１６８】
また、上述した第１～第５の実施形態に係る音声再生装置は、ＬＳＩなどの集積回路や、専用の信号処理回路を用いて１チップ化したものによって実現されてもよい。また上述した第１～第５の実施形態に係る音声再生装置は、音声再生装置を構成する各構成部の機能に相当するものをそれぞれチップ化したものによって実現されてもよい。なお、ここでは、ＬＳＩとしたが、集積度の違いにより、ＩＣ、システムＬＳＩ、スーパーＬＳＩ、ウルトラＬＳＩと呼称されることもある。また集積回路化の手法は、ＬＳＩに限るものではなく、専用回路又は汎用プロセッサで実現してもよい。ＬＳＩ製造後に、プログラムすることが可能なＦＰＧＡ（Ｆｉｅｌｄ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｇａｔｅ　Ａｒｒａｙ）や、ＬＳＩ内部の回路セルの接続や設定を再構成可能なリコンフィギュラブル・プロセッサを利用してもよい。さらには、半導体技術の進歩又は派生する別技術によりＬＳＩに置き換わる集積回路化の技術が登場すれば、当然、その技術を用いて機能ブロックの集積化を行ってもよい。
【産業上の利用可能性】
【０１６９】
本発明に係る音声再生装置は、目標時間を達成しつつ、入力されるオーディオ信号に応じた適切な速度変換を行うことが可能なハードディスクレコーダーやＤＶＤレコーダー等のＡＶコンテンツ視聴用機器、パソコンや携帯電話等のモバイル機器上で動作するアプリケーション等に有用である。また、視聴用途だけではなく、学習コンテンツ再生システム等において内容の理解を容易にするための用途や、セキュリティカメラで撮影された映像等の長時間のコンテンツについて概要の把握を容易にするための用途等にも有用である。
【図面の簡単な説明】
【０１７０】
【図１】第１の実施形態に係る音声再生装置の構成例を示すブロック図
【図２】ジャンル別の音声含有率を示した図
【図３】各ジャンルの音声含有率の平均と標準偏差とを示した図
【図４】５種類の算出パターンを示した図
【図５】速度比算出分布の一例を示す図
【図６】音声含有率が０．５のときの目標圧伸比、音声区間の平均速度比、非音声区間の平均速度比を示す図
【図７】音声及び非音声区間の速度比変化を示した模式図
【図８】第１の実施形態に係る音声再生装置の処理の流れを示すフローチャート
【図９】ニュース番組に含まれる音声区間長とその頻度を示した図
【図１０】野球番組に含まれる音声区間長とその頻度を示した図

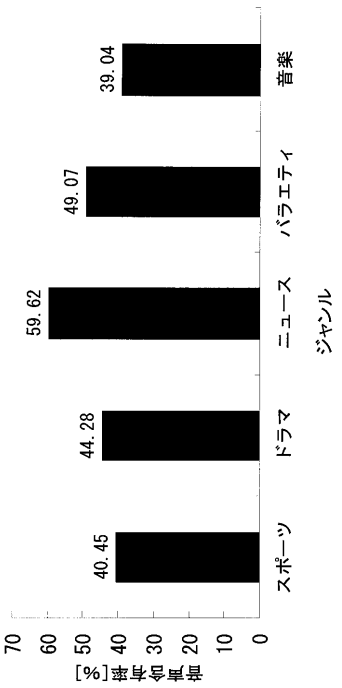
【図１１】音声区間の圧伸比の変化を示した図
【図１２】２段階の変換速度比を算出した場合を示す図
【図１３】ドキュメンタリーなどの静止画像が多いジャンルについての速度比算出分布の例を示す図
【図１４】図１３に示す速度比算出分布において、音声含有率が０．５のときの目標圧伸比、音声区間の平均速度比、非音声区間の平均速度比を示す図
【図１５】スポーツなど動きの激しいシーンが多いジャンルについての速度比算出分布の例を示す図
【図１６】図１５に示す速度比算出分布において、音声含有率が０．５のときの目標圧伸比、音声区間の平均速度比、非音声区間の平均速度比を示す図
【図１７】第２の実施形態に係る音声再生装置の構成例を示すブロック図
【図１８】式（１１）で示される予測音声含有率Ｙ（ｚ）を求める方法を模式的に示す図
【図１９】音声含有率Ｘ（ｚ）と予測音声含有率Ｙ（ｚ）の算出結果の一例を示す図
【図２０】式（１４）に基づいて予測した音声区間長を示す図
【図２１】第２の実施形態に係る音声再生装置の処理の流れを示すフローチャート
【図２２】第３の実施形態に係る音声再生装置の構成例を示すブロック図
【図２３】コンテンツ毎の長期音声含有率の時間変化を示す図
【図２４】コンテンツ毎の予測音声含有率を示す図
【図２５】第３の実施形態に係る速度比条件設定部１４の処理を示すフローチャート
【図２６】音声区間長の実測値と、直前の音声区間長の実測値と、予測音声区間長の分布を示した図
【図２７】第４の実施形態に係る音声再生装置の構成例を示すブロック図
【図２８】第４の実施形態に係る音声再生装置の処理の流れを示すフローチャート
【図２９】第５の実施形態に係る音声再生装置の構成例を示すブロック図
【図３０】音楽含有率と音楽速度比との対応を示したテーブルの例を示す図
【図３１】第５の実施形態に係る音声再生装置の処理の流れを示すフローチャート
【図３２】音声再生装置がコンピュータシステム５０によって実現される構成例を示すブロック図
【図３３】従来の音声再生装置の構成を示したブロック図
【符号の説明】
【０１７１】
１１　音声非音声判別部
１２　蓄積部
１３　音声含有率算出部
１４　速度比条件設定部
１５　音声区間長算出部
１６　速度比決定部
１７　速度変換部
１８　音声含有率予測部
１９　音声区間長予測部
２０　圧伸比算出部
２１　統計量算出部
２２　一時蓄積部
２３　音判別部
２４　特定イベント含有率算出部
５０　コンピュータシステム
５１　ＣＰＵ
５２　メモリ
５３　ハードディスク
５４　ディスクドライブ装置

５５　　モニタ
５６　　スピーカ
５７　　入力装置

【図１】



【図２】



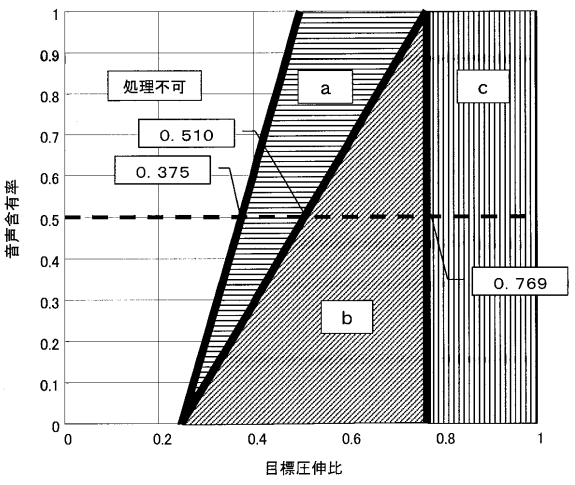
【図３】

スポーツ	音声含有率	ドラマ・映画・アニメ	音声含有率	ニュース	音声含有率
平均	40.4	平均	44.3	平均	59.6
標準偏差	5.3	標準偏差	16.2	標準偏差	8.2

【図４】

平均速度比算出パターン	a	b	c	d	e
音声区間	$Vm1=An\cdot S/(An\cdot E-1+S)$	$Vm1=Bs$	$Vm1=1/E$	$Vm1=Dn\cdot S/(Dn\cdot E-1+S)$	$Vm1=Es$
非音声区間	$Vm2=An$	$Vm2=(1-S)\cdot Bs/(Bs\cdot E-S)$	$Vm2=1/E$	$Vm2=Dn$	$Vm2=(1-S)\cdot Es/(Es\cdot E-S)$
算出条件	$Vm1 \leqq Vm2$ $Vm2=An(一定)$	$Vm1 \leqq Vm2$ $Vm1=Bs(一定)$	$Vm1=Vm2$	$Vm1 \geqq Vm2$ $Vm2=Dn(一定)$	$Vm1 \geqq Vm2$ $Vm1=Es(一定)$

【図５】

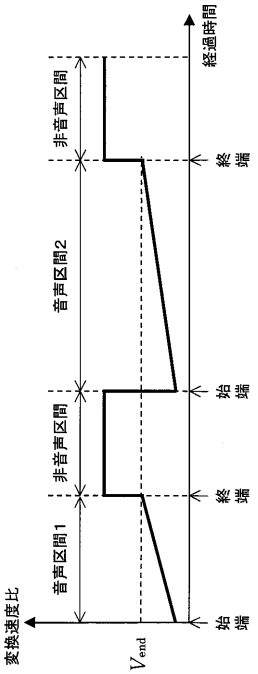


【図６】

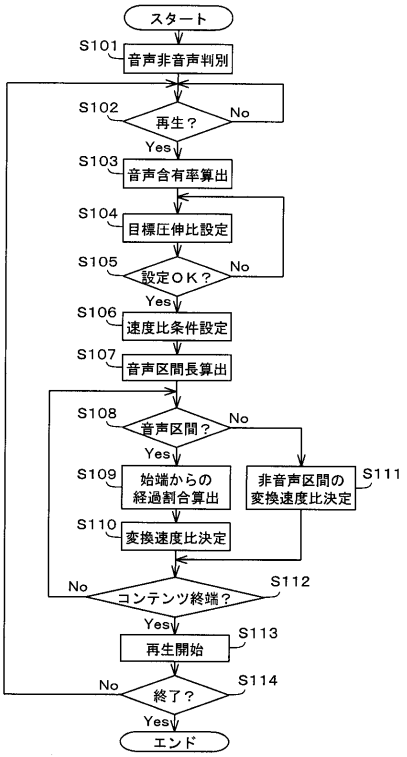
音声含有率0.5のときの平均速度比　Ar=4　Bs=1.3

目標圧伸比		0.1	0.3	0.4	0.5	0.6	0.7	0.9	1
算出パターン		不可	不可	a	a	b	b	c	c
平均速度比	音声区間	不可	不可	1.818	1.333	1.300	1.300	1.111	1.000
	非音声区間	不可	不可	4.000	4.000	2.321	1.585	1.111	1.000

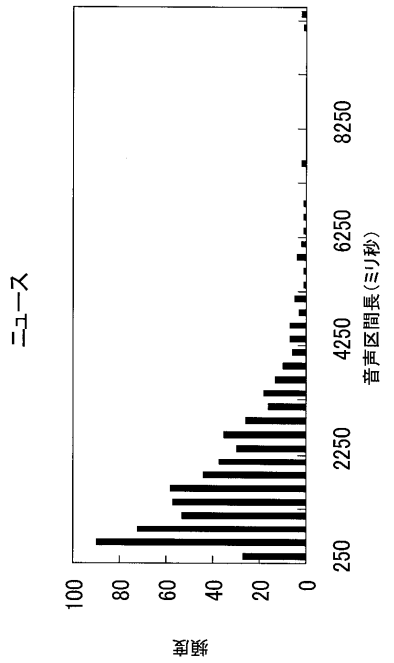
【図７】


経過時間
非音声区間
音声区間2
非音声区間
音声区間1
終端
始端
終端
始端
変換速度比
V_{end}

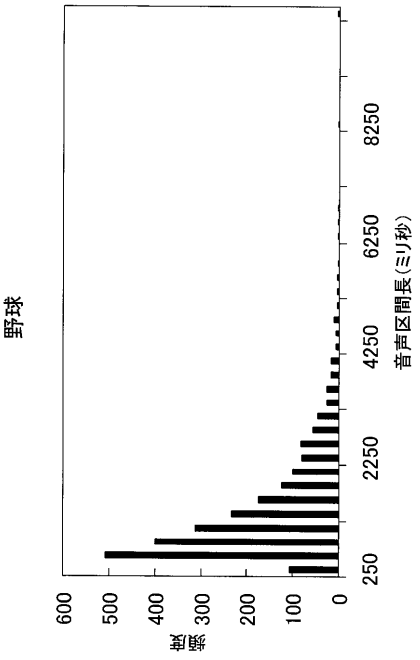
【図８】


スタート
S101
音声非音声判別
S102
再生？
No
Yes
S103
音声含有率算出
S104
目標圧伸比設定
S105
設定ＯＫ？
No
Yes
S106
速度比条件設定
S107
音声区間長算出
S108
音声区間？
No
Yes
S109
始端からの経過割合算出
S110
変換速度比決定
S111
非音声区間の変換速度比決定
S112
No
コンテンツ終端？
Yes
S113
再生開始
No
S114
終了？
Yes
エンド

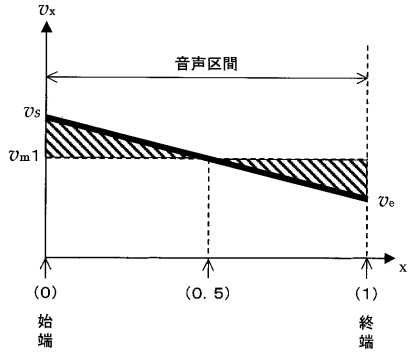
【図９】


ニュース
頻度
音声区間長(ミリ秒)
100
80
60
40
20
0
250
2250
4250
6250
8250

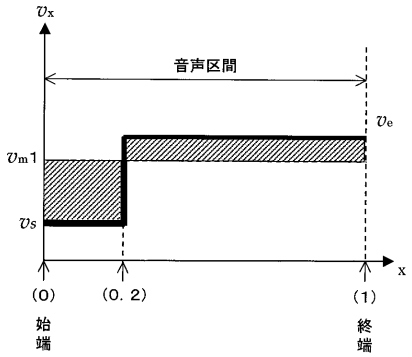
【図１０】


野球
頻度
音声区間長(ミリ秒)
600
500
400
300
200
100
0
250
2250
4250
6250
8250

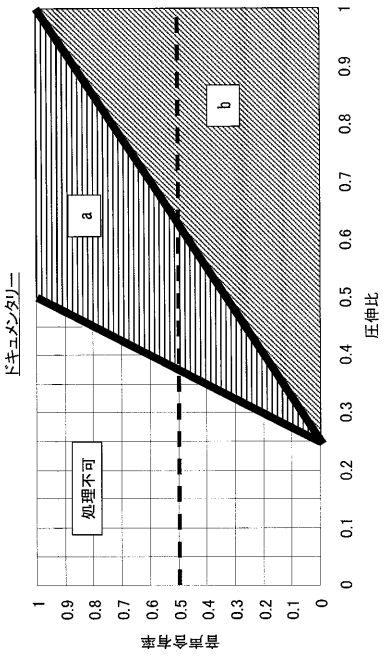
【図１１】



【図１２】



【図１３】



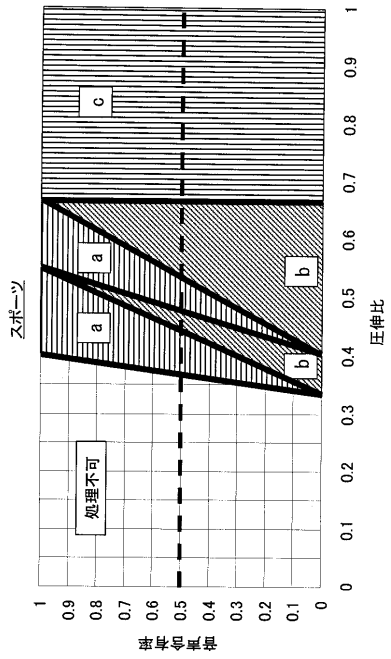
【図１４】

ドキュメンタリー

音声含有率0.50のときの平均速度比

目標圧伸比		0.1	0.3	Ar= 4 0.4	0.5	Bs= 1 0.6	0.7	0.9	1
算出パターン		不可	不可	a	a	a	b	b	b
平均速度比	音声区間	不可	不可	1.818	1.333	1.053	1.000	1.000	1.000
	非音声区間	不可	不可	4.000	4.000	4.000	2.500	1.250	1.000

【図１５】

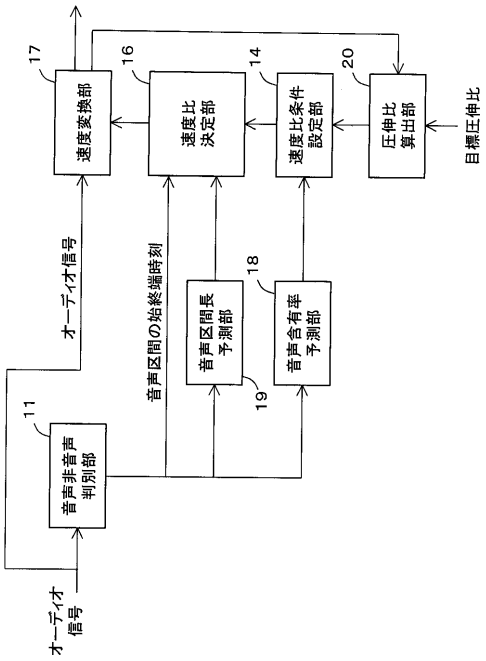


【図１６】

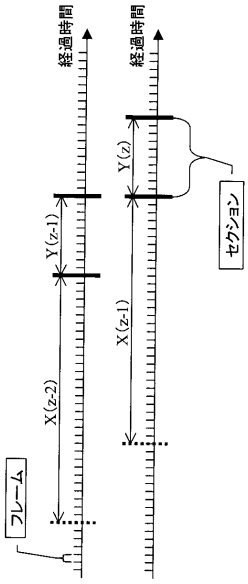
スポーツ

音声含有率0.5のときの平均速度比			Ar= 3		2.5		Bs= 1.8		1.5	
目標圧伸比		0.1	0.3	0.4	0.5	0.6	0.7	0.9	1	
算出パターン		不可	不可	a(1)	a(2)	b(2)	c	c	c	
平均速度比	音声区間	不可	不可	2.143	1.667	1.500	1.429	1.111	1.000	
	非音声区間	不可	不可	3.000	2.500	1.875	1.429	1.111	1.000	

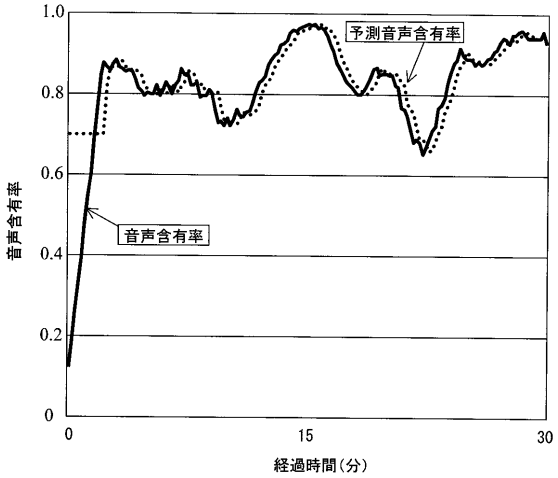
【図１７】



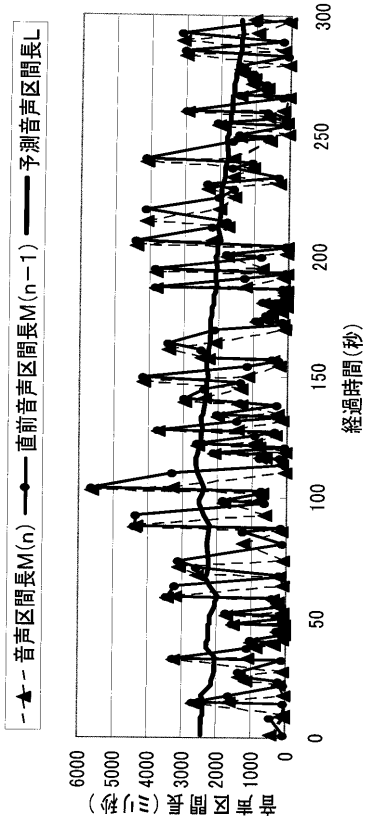
【図１８】



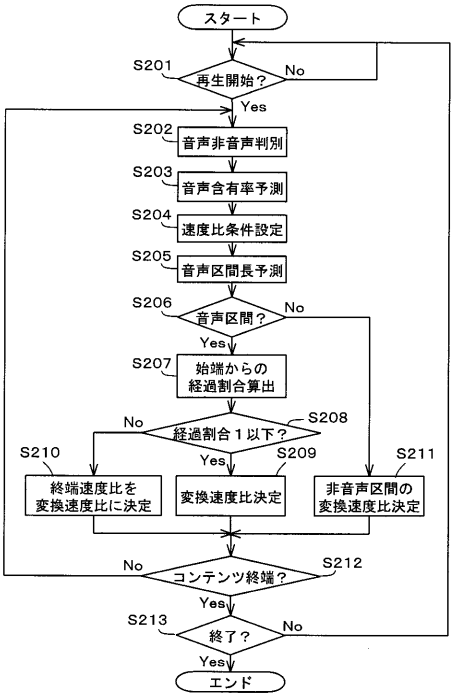
【図１９】



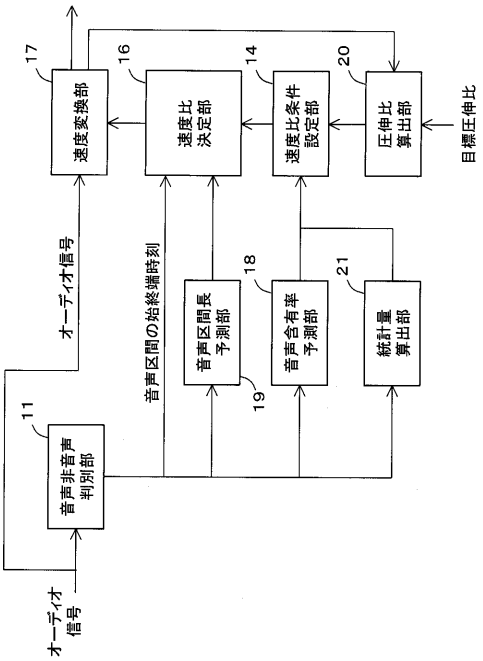
【図２０】



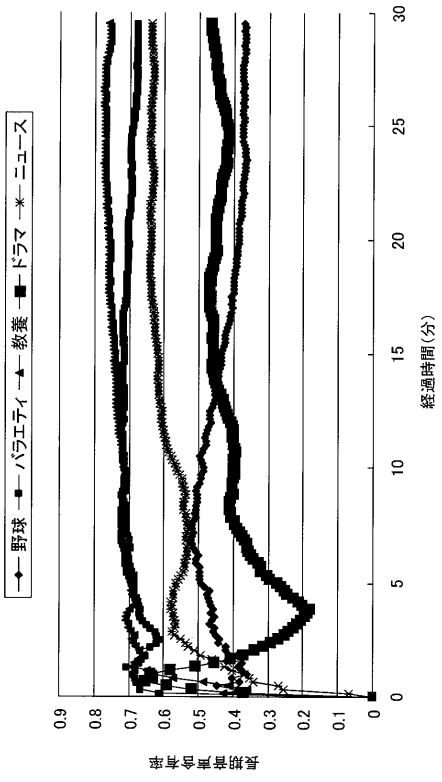
【図２１】



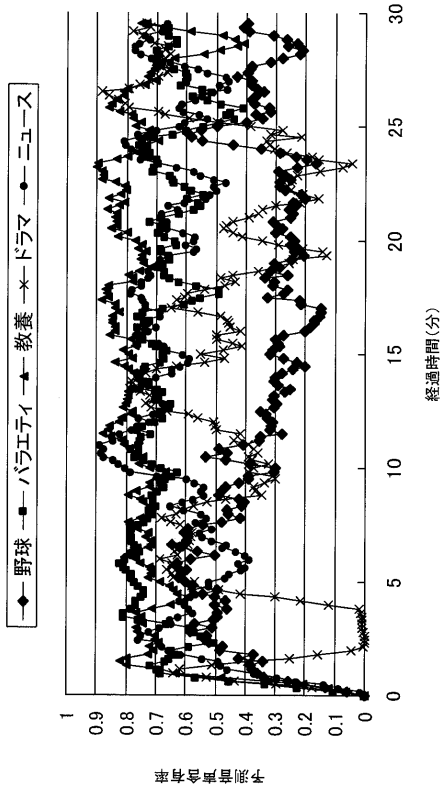
【図２２】



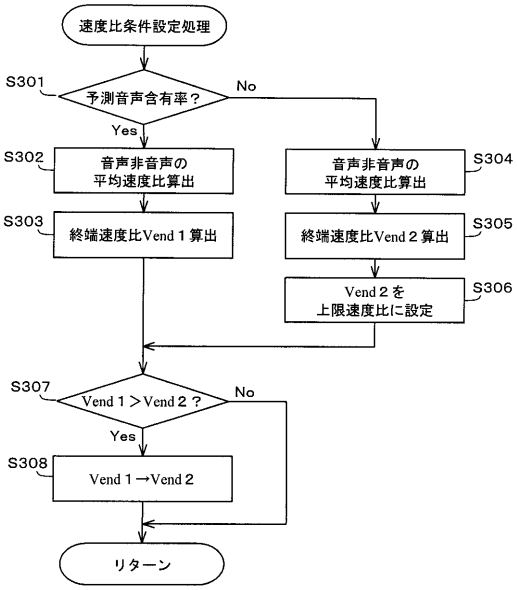
【図２３】



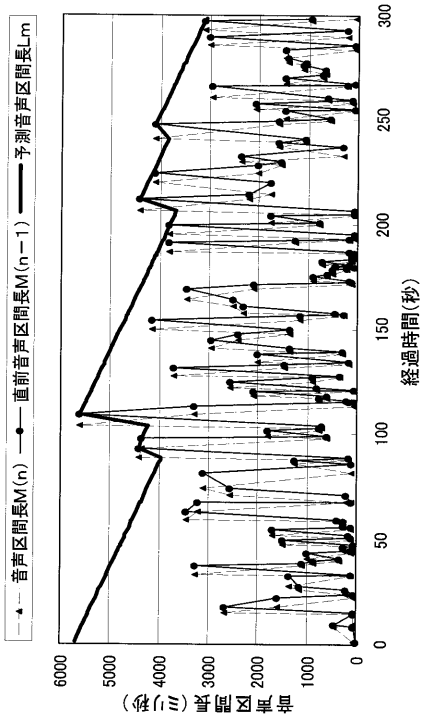
【図２４】



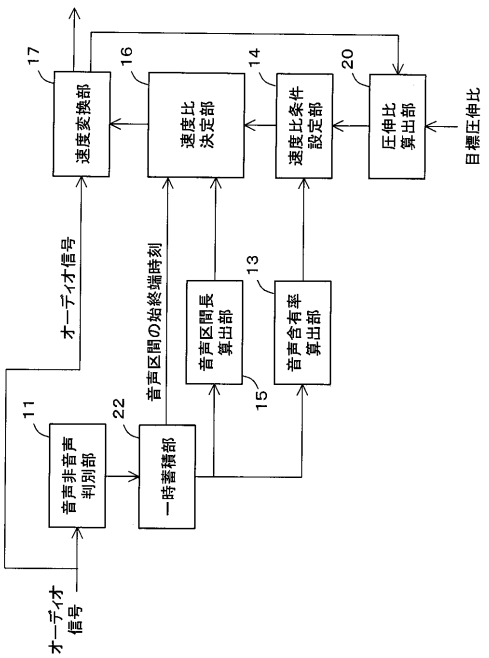
【図２５】



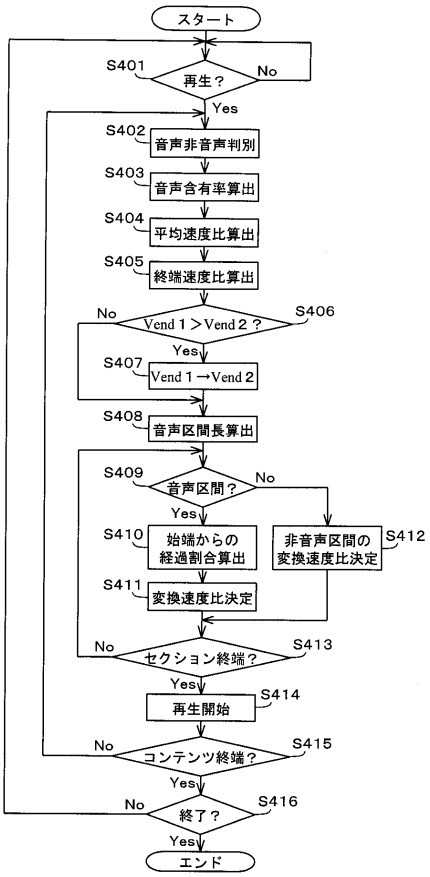
【図２６】



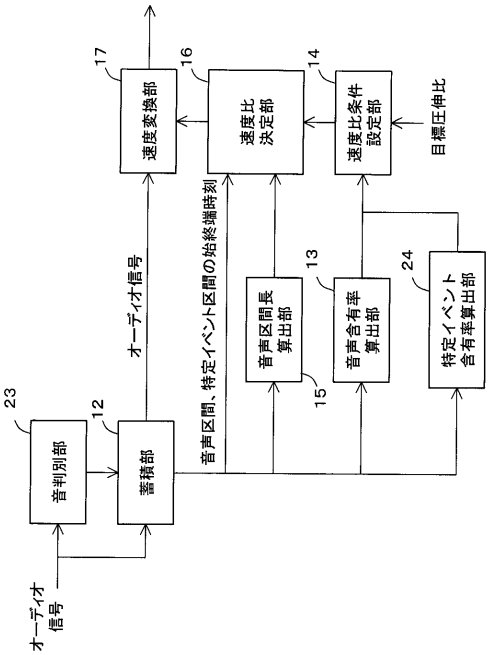
【図２７】



【図２８】



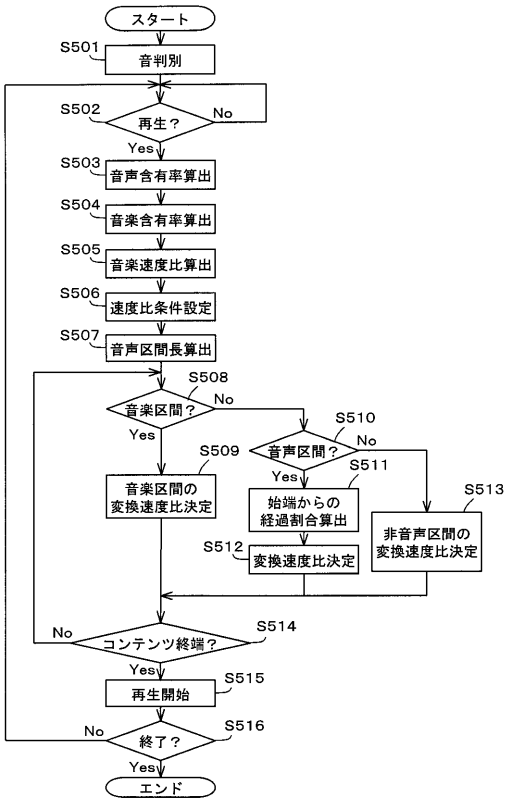
【図２９】



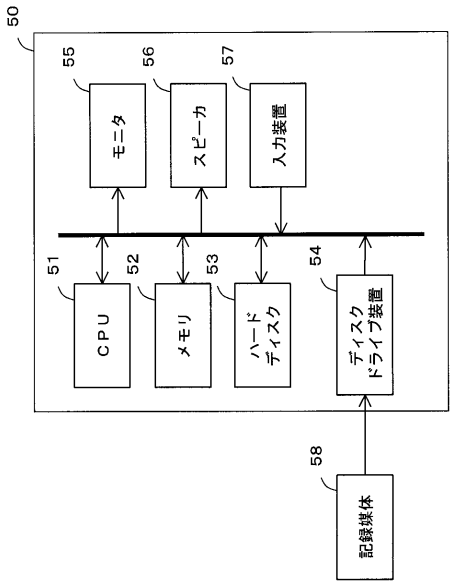
【図３０】

音楽含有率	0%～30%	～40%	～50%	～60%	～70%	～100%
音楽速度比	1	1.1	1.2	1.3	1.5	2

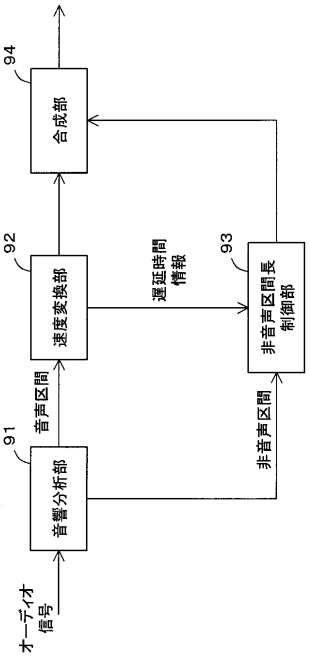
【図３１】



【図３２】


50
55
モニタ
56
スピーカ
57
入力装置
51
CPU
52
メモリ
53
ハード
ディスク
54
ディスク
ドライブ装置
58
記録媒体

【図３３】


94
合成部
92
速度変換部
遅延時間
情報
93
非音声区間長
制御部
音声区間
91
音響分析部
非音声区間
オーディオ
信号

フロントページの続き

審査官　菊池　智紀

(56)参考文献　特開平０８－２５４９９２（ＪＰ，Ａ）
特開平０４－３６７８９８（ＪＰ，Ａ）
特開平０６－３３７６９６（ＪＰ，Ａ）
特開平０８－３２８５８６（ＪＰ，Ａ）
特開２００１－１８４１００（ＪＰ，Ａ）
特開平０９－１４６５８７（ＪＰ，Ａ）
特開２００５－１４８３０７（ＪＰ，Ａ）
特開２００５－２６６５７１（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１９／００－２１／０６
Ｇ１１Ｂ　２０／１０
ＪＳＴＰｌｕｓ（ＪＤｒｅａｍＩＩ）