

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7200483号
(P7200483)

(45)発行日 令和5年1月10日(2023.1.10)

(24)登録日 令和4年12月26日(2022.12.26)

(51)国際特許分類 F I
 G 1 0 L 21/013(2013.01) G 1 0 L 21/013
 G 1 0 L 21/01 (2013.01) G 1 0 L 21/01
 G 1 0 L 21/043(2013.01) G 1 0 L 21/043

請求項の数 10 (全20頁)

(21)出願番号	特願2018-43118(P2018-43118)	(73)特許権者	000004075 ヤマハ株式会社 静岡県浜松市中区中沢町10番1号
(22)出願日	平成30年3月9日(2018.3.9)	(74)代理人	100125689 弁理士 大林 章
(65)公開番号	特開2019-159014(P2019-159014 A)	(74)代理人	100128598 弁理士 高田 聖一
(43)公開日	令和1年9月19日(2019.9.19)	(74)代理人	100121108 弁理士 高橋 太郎
審査請求日	令和3年1月21日(2021.1.21)	(72)発明者	大道 竜之介 静岡県浜松市中区中沢町10番1号 ヤ マハ株式会社内
		(72)発明者	嘉山 啓 静岡県浜松市中区中沢町10番1号 ヤ マハ株式会社内

最終頁に続く

(54)【発明の名称】 音声処理方法、音声処理装置およびプログラム

(57)【特許請求の範囲】

【請求項1】

歌唱音声を表す第1音信号において基本周波数およびスペクトル形状が時間的に安定している第1定常期間の終点と、前記歌唱音声とは音響特性が相違する参照音声を表す第2音信号において基本周波数およびスペクトル形状が時間的に安定している第2定常期間の終点とが時間軸上で一致するように、前記第1定常期間と前記第2定常期間との時間軸上の位置を調整した状態において、前記第1音信号のうち前記第1定常期間の終点に対して特定の時間だけ前方の時点から前記歌唱音声が消音する時点までの処理期間を、前記第2音信号のうち前記第2定常期間の終点に対して前記特定の時間だけ前方の時点から前記参照音声が消音する時点までの表現期間の時間長に伸長し、

10

前記第2音信号の前記表現期間における声質に関する音響特性を、前記伸長後の前記第1音信号における前記処理期間に付加する、

コンピュータにより実現される音声処理方法。

【請求項2】

前記処理期間の伸長においては、当該処理期間の終点に近い位置ほど伸長の度合いが小さくなるように前記処理期間を伸長する

請求項1の音声処理方法。

【請求項3】

歌唱音声を表す第1音信号のうち当該歌唱音声を開始する時点から、前記第1音信号において基本周波数およびスペクトル形状が時間的に安定している第1定常期間の始点まで

20

の処理期間を、前記歌唱音声とは音響特性が相違する参照音声を表す第 2 音信号のうち前記参照音声が開始する時点から、前記第 2 音信号において基本周波数およびスペクトル形状が時間的に安定している第 2 定常期間の始点までの表現期間の時間長に伸長し、

前記第 2 音信号の前記表現期間における声質に関する音響特性を、前記伸長後の前記第 1 音信号における前記処理期間に付加する、

コンピュータにより実現される音声処理方法。

【請求項 4】

前記処理期間の伸長においては、前記処理期間の始点に近い位置ほど伸長の度合いが小さくなるように前記処理期間を伸長する

請求項 3 の音声処理方法。

10

【請求項 5】

歌唱音声を表す第 1 音信号において基本周波数およびスペクトル形状が時間的に安定している第 1 定常期間の終点と、前記歌唱音声とは音響特性が相違する参照音声を表す第 2 音信号において基本周波数およびスペクトル形状が時間的に安定している第 2 定常期間の終点とが時間軸上で一致するように、前記第 1 定常期間と前記第 2 定常期間との時間軸上の位置を調整した状態において、前記第 1 音信号のうち前記第 1 定常期間の終点に対して特定の時間だけ前方の時点から前記歌唱音声が消音する時点までの処理期間を、前記第 2 音信号のうち前記第 2 定常期間の終点に対して前記特定の時間だけ前方の時点から前記参照音声が消音する時点までの表現期間の時間長に伸長し、前記第 2 音信号の前記表現期間における声質に関する音響特性を、前記伸長後の前記第 1 音信号における前記処理期間に付加する合成処理部

20

を具備する音声処理装置。

【請求項 6】

前記合成処理部は、当該処理期間の終点に近い位置ほど伸長の度合いが小さくなるように前記処理期間を伸長する

請求項 5 の音声処理装置。

【請求項 7】

歌唱音声を表す第 1 音信号のうち当該歌唱音声が開始する時点から、前記第 1 音信号において基本周波数およびスペクトル形状が時間的に安定している第 1 定常期間の始点までの処理期間を、前記歌唱音声とは音響特性が相違する参照音声を表す第 2 音信号のうち前記参照音声が開始する時点から、前記第 2 音信号において基本周波数およびスペクトル形状が時間的に安定している第 2 定常期間の始点までの表現期間の時間長に伸長し、前記第 2 音信号の前記表現期間における声質に関する音響特性を、前記伸長後の前記第 1 音信号における前記処理期間に付加する合成処理部

30

を具備する音声処理装置。

【請求項 8】

前記合成処理部は、前記処理期間の始点に近い位置ほど伸長の度合いが小さくなるように前記処理期間を伸長する

請求項 7 の音声処理装置。

【請求項 9】

40

歌唱音声を表す第 1 音信号において基本周波数およびスペクトル形状が時間的に安定している第 1 定常期間の終点と、前記歌唱音声とは音響特性が相違する参照音声を表す第 2 音信号において基本周波数およびスペクトル形状が時間的に安定している第 2 定常期間の終点とが時間軸上で一致するように、前記第 1 定常期間と前記第 2 定常期間との時間軸上の位置を調整した状態において、前記第 1 音信号のうち前記第 1 定常期間の終点に対して特定の時間だけ前方の時点から前記歌唱音声が消音する時点までの処理期間を、前記第 2 音信号のうち前記第 2 定常期間の終点に対して前記特定の時間だけ前方の時点から前記参照音声が消音する時点までの表現期間の時間長に伸長し、前記第 2 音信号の前記表現期間における声質に関する音響特性を、前記伸長後の前記第 1 音信号における前記処理期間に付加する合成処理部、

50

としてコンピュータを機能させるプログラム。

【請求項 10】

歌唱音声を表す第 1 音信号のうち当該歌唱音声を開始する時点から、前記第 1 音信号において基本周波数およびスペクトル形状が時間的に安定している第 1 定常期間の始点までの処理期間を、前記歌唱音声とは音響特性が相違する参照音声を表す第 2 音信号のうち前記参照音声を開始する時点から、前記第 2 音信号において基本周波数およびスペクトル形状が時間的に安定している第 2 定常期間の始点までの表現期間の時間長に伸長し、前記第 2 音信号の前記表現期間における声質に関する音響特性を、前記伸長後の前記第 1 音信号における前記処理期間に付加する合成処理部、

としてコンピュータを機能させるプログラム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声を表す音声信号を処理する技術に関する。

【背景技術】

【0002】

歌唱表現等の音声表現を音声に付加する各種の技術が従来から提案されている。例えば特許文献 1 には、音声信号の各調波成分を周波数領域で移動させることにより、当該音声信号が表す音声を、濁声または嗄声等の特徴的な声質の音声に変換する技術が開示されている。

20

【先行技術文献】

【特許文献】

【0003】

【文献】特開 2014 - 2338 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかし、特許文献 1 の技術においては、聴感的に自然な音声を生成するという観点から更なる改善の余地がある。以上の事情を考慮して、本発明は、聴感的に自然な音声を合成することを目的とする。

30

【課題を解決するための手段】

【0005】

以上の課題を解決するために、本発明の好適な態様に係る音声処理方法は、歌唱音声を表す第 1 音信号のうちの処理期間を、前記歌唱音声とは音響特性が相違する参照音声を表す前記第 2 音信号において前記第 1 音信号の変形に適用されるべき表現期間の時間長に応じて伸長し、前記処理期間の伸長後の前記第 1 音信号を、前記第 2 音信号の前記表現期間に応じて変形する。

【0006】

以上の課題を解決するために、本発明の好適な態様に係る音声処理装置は、歌唱音声を表す第 1 音信号のうちの処理期間を、前記歌唱音声とは音響特性が相違する参照音声を表す前記第 2 音信号において前記第 1 音信号の変形に適用されるべき表現期間の時間長に応じて伸長し、前記処理期間の伸長後の前記第 1 音信号を、前記第 2 音信号の前記表現期間に応じて変形する合成処理部を具備する。

40

【図面の簡単な説明】

【0007】

【図 1】本発明の実施形態に係る音処理装置の構成を例示するブロック図である。

【図 2】音処理装置の機能的な構成を例示するブロック図である。

【図 3】第 1 音信号における定常期間の説明図である。

【図 4】信号解析処理の具体的な手順を例示するフローチャートである。

【図 5】歌唱音声の発音が始まった直後における基本周波数の時間変化である。

50

【図 6】歌唱音声の発音が終了する直前における基本周波数の時間変化である。

【図 7】リリース処理の具体的な手順を例示するフローチャートである。

【図 8】リリース処理の説明図である。

【図 9】スペクトル包絡概形の説明図である。

【図 10】アタック処理の具体的な手順を例示するフローチャートである。

【図 11】アタック処理の説明図である。

【発明を実施するための形態】

【0008】

図 1 は、本発明の好適な形態に係る音処理装置 100 の構成を例示するブロック図である。本実施形態の音処理装置 100 は、利用者が楽曲を歌唱した音声（以下「歌唱音声」という）に対して各種の音表現を付加する信号処理装置である。音表現は、歌唱音声（第 1 音の例示）に対して付加される音響特性である。楽曲の歌唱に着目すると、音表現は、音声の発音（すなわち歌唱）に関する音楽的な表現または表情である。具体的には、ボーカルフライ、唸り声、または囁れ声のような歌唱表現が、音表現の好適例である。なお、音表現は、声質とも換言される。

10

【0009】

音表現は、発音の開始の直後に音量が増加していく部分（以下「アタック部」という）と、発音の終了の直前に音量が減少していく部分（以下「リリース部」という）とにおいて特に顕著となる。以上の傾向を考慮して、本実施形態では、歌唱音声のうち特にアタック部およびリリース部に対して音表現を付加する。

20

【0010】

図 1 に例示される通り、音処理装置 100 は、制御装置 11 と記憶装置 12 と操作装置 13 と放音装置 14 とを具備するコンピュータシステムで実現される。例えば携帯電話機もしくはスマートフォン等の可搬型の情報端末、またはパーソナルコンピュータ等の可搬型または据置型の情報端末が、音処理装置 100 として好適に利用される。操作装置 13 は、利用者からの指示を受付ける入力機器である。例えば、利用者が操作する複数の操作子、または利用者による接触を検知するタッチパネルが、操作装置 13 として好適に利用される。

【0011】

制御装置 11 は、例えば CPU (Central Processing Unit) 等の処理回路であり、各種の演算処理および制御処理を実行する。本実施形態の制御装置 11 は、歌唱音声に音表現を付与した音声（以下「変形音」という）を表す第 3 音信号 Y を生成する。放音装置 14 は、例えばスピーカまたはヘッドホンであり、制御装置 11 が生成した第 3 音信号 Y が表す変形音を放音する。なお、制御装置 11 が生成した第 3 音信号 Y をデジタルからアナログに変換する D/A 変換器の図示は便宜的に省略した。なお、音処理装置 100 が放音装置 14 を具備する構成を図 1 では例示したが、音処理装置 100 とは別体の放音装置 14 を音処理装置 100 に有線または無線で接続してもよい。

30

【0012】

記憶装置 12 は、例えば磁気記録媒体または半導体記録媒体等の公知の記録媒体で構成されたメモリであり、制御装置 11 が実行するプログラムと制御装置 11 が使用する各種のデータとを記憶する。なお、複数種の記録媒体の組合せにより記憶装置 12 を構成してもよい。また、音処理装置 100 とは別体の記憶装置 12（例えばクラウドストレージ）を用意し、制御装置 11 が通信網を介して記憶装置 12 に対する書込および読出を実行してもよい。すなわち、記憶装置 12 を音処理装置 100 から省略してもよい。

40

【0013】

本実施形態の記憶装置 12 は、第 1 音信号 X1 と第 2 音信号 X2 とを記憶する。第 1 音信号 X1 は、音処理装置 100 の利用者が楽曲を歌唱した歌唱音声を表す音響信号である。第 2 音信号 X2 は、利用者以外の歌唱者（例えば歌手）が音表現を付加して歌唱した音声（以下「参照音声」という）を表す音響信号である。第 1 音信号 X1 と第 2 音信号 X2 とでは音響特性（例えば声質）が相違する。本実施形態の音処理装置 100 は、第 2 音信号 X

50

2が表す参照音声（第2音の例示）の音表現を、第1音信号X1が表す歌唱音声に付加することで、変形音の第3音信号Yを生成する。なお、歌唱音声と参照音声との間で楽曲の異同は不問である。なお、以上の説明では歌唱音声の発声者と参照音声の発声者とが別人である場合を想定したが、歌唱音声の発声者と参照音声の発声者とは同一人でもよい。例えば、歌唱音声は、音表現を付加せずに利用者が歌唱した音声であり、参照音声は、当該利用者が歌唱表現を付加した音声である。

【0014】

図2は、制御装置11の機能的な構成を例示するブロック図である。図2に例示される通り、制御装置11は、記憶装置12に記憶されたプログラムを実行することで、第1音信号X1と第2音信号X2とから第3音信号Yを生成するための複数の機能（信号解析部21および合成処理部22）を実現する。なお、相互に別体で構成された複数の装置で制御装置11の機能を実現してもよいし、制御装置11の機能の一部または全部を専用の電子回路で実現してもよい。

10

【0015】

信号解析部21は、第1音信号X1の解析により解析データD1を生成し、第2音信号X2の解析により解析データD2を生成する。信号解析部21が生成した解析データD1および解析データD2は記憶装置12に格納される。

【0016】

解析データD1は、第1音信号X1における複数の定常期間Q1を表すデータである。図3に例示される通り、解析データD1が示す各定常期間Q1は、第1音信号X1のうち基本周波数f1とスペクトル形状とが時間的に安定している可変長の期間である。解析データD1は、各定常期間Q1の始点の時刻（以下「始点時刻」という）T1_Sと終点の時刻（以下「終点時刻」という）T1_Eとを指定する。なお、楽曲内で相前後する2個の音符の間では、基本周波数f1またはスペクトル形状（すなわち音韻）が変化する場合が多い。したがって、各定常期間Q1は、楽曲内の1個の音符に相当する期間である可能性が高い。

20

【0017】

同様に、解析データD2は、第2音信号X2における複数の定常期間Q2を表すデータである。各定常期間Q2は、第2音信号X2のうち基本周波数f2とスペクトル形状とが時間的に安定している可変長の期間である。解析データD2は、各定常期間Q2の始点時刻T2_Sと終点時刻T2_Eとを指定する。定常期間Q1と同様に、各定常期間Q2は、楽曲内の1個の音符に相当する期間である可能性が高い。

30

【0018】

図4は、信号解析部21が第1音信号X1を解析する処理（以下「信号解析処理」という）S0のフローチャートである。例えば操作装置13に対する利用者からの指示を契機として図4の信号解析処理S0が開始される。図4に例示される通り、信号解析部21は、時間軸上の複数の単位期間（フレーム）の各々について第1音信号X1の基本周波数f1を算定する（S01）。基本周波数f1の算定には公知の技術が任意に採用される。各単位期間は、定常期間Q1に想定される時間長と比較して十分に短い期間である。

【0019】

信号解析部21は、第1音信号X1のスペクトル形状を表すメルケプストラムM1を単位期間毎に算定する（S02）。メルケプストラムM1は、第1音信号X1の周波数スペクトルの包絡線を表す複数の係数で表現される。メルケプストラムM1は、歌唱音声の音韻を表す特徴量とも表現される。メルケプストラムM1の算定には公知の技術が任意に採用される。なお、第1音信号X1のスペクトル形状を表す特徴量として、メルケプストラムM1の代わりにMFCC（Mel-Frequency Cepstrum Coefficients）を算定してもよい。

40

【0020】

信号解析部21は、第1音信号X1が表す歌唱音声の有声性を単位期間毎に推定する（S03）。すなわち、歌唱音声が有声音および無声音の何れに該当するかが判定される。有声性（有声/無声）の推定には公知の技術が任意に採用される。なお、基本周波数f1の算定（S01）とメルケプストラムM1の算定（S02）と有声性の推定（S03）とについ

50

て順序は任意であり、以上に例示した順序には限定されない。

【 0 0 2 1 】

信号解析部 2 1 は、基本周波数 f_1 の時間的な変化の度合を示す第 1 指標 1 を単位期間毎に算定する (S 0 4)。例えば相前後する 2 個の単位期間の間における基本周波数 f_1 の差分が第 1 指標 1 として算定される。基本周波数 f_1 の時間的な変化が顕著であるほど第 1 指標 1 は大きい数値となる。

【 0 0 2 2 】

信号解析部 2 1 は、メルケプストラム M_1 の時間的な変化の度合を示す第 2 指標 2 を単位期間毎に算定する (S 0 5)。例えば、相前後する 2 個の単位期間の間においてメルケプストラム M_1 の係数毎の差分を複数の係数について合成 (例えば加算または平均) した数値が、第 2 指標 2 として好適である。歌唱音声のスペクトル形状の時間的な変化が顕著であるほど第 2 指標 2 は大きい数値となる。例えば歌唱音声の音韻が変化する時点の付近では、第 2 指標 2 は大きい数値となる。

10

【 0 0 2 3 】

信号解析部 2 1 は、第 1 指標 1 および第 2 指標 2 に応じた変動指標 を単位期間毎に算定する (S 0 6)。例えば、第 1 指標 1 と第 2 指標 2 との加重和が変動指標 として単位期間毎に算定される。第 1 指標 1 および第 2 指標 2 の各々の加重値は、所定の固定値、または操作装置 1 3 に対する利用者からの指示に応じた可変値に設定される。以上の説明から理解される通り、第 1 音信号 X_1 の基本周波数 f_1 またはメルケプストラム M_1 (すなわちスペクトル形状) の時間的な変動が大きいほど、変動指標 は大きい数値になるという傾向がある。

20

【 0 0 2 4 】

信号解析部 2 1 は、第 1 音信号 X_1 における複数の定常期間 Q_1 を特定する (S 0 7)。本実施形態の信号解析部 2 1 は、歌唱音声の有声性の推定の結果 (S 0 3) と変動指標 とに応じて定常期間 Q_1 を特定する。具体的には、信号解析部 2 1 は、歌唱音声の有声音であると推定され、かつ、変動指標 が所定の閾値を下回る一連の単位期間の集合を定常期間 Q_1 として画定する。歌唱音声が無声音であると推定された単位期間、または、変動指標 が閾値を上回る単位期間は、定常期間 Q_1 から除外される。以上の手順により第 1 音信号 X_1 の各定常期間 Q_1 を画定すると、信号解析部 2 1 は、各定常期間 Q_1 の始点時刻 T_{1_S} と終点時刻 T_{1_E} とを指定する解析データ D_1 を記憶装置 1 2 に格納する (S 0 8)。

30

【 0 0 2 5 】

信号解析部 2 1 は、以上に説明した信号解析処理 S_0 を、参照音声を表す第 2 音信号 X_2 についても実行することで解析データ D_2 を生成する。具体的には、信号解析部 2 1 は、第 2 音信号 X_2 の単位期間毎に、基本周波数 f_2 の算定 (S 0 1) とメルケプストラム M_2 の算定 (S 0 2) と有声性 (有声 / 無声) の推定 (S 0 3) とを実行する。信号解析部 2 1 は、基本周波数 f_2 の時間的な変化の度合を示す第 1 指標 1 と、メルケプストラム M_2 の時間的な変化の度合を示す第 2 指標 2 とに応じた変動指標 を算定する (S 0 4 - S 0 6)。そして、信号解析部 2 1 は、参照音声の有声性の推定の結果 (S 0 3) と変動指標 とに応じて第 2 音信号 X_2 の各定常期間 Q_2 を特定する (S 0 7)。信号解析部 2 1 は、各定常期間 Q_2 の始点時刻 T_{2_S} と終点時刻 T_{2_E} とを指定する解析データ D_2 を記憶装置 1 2 に格納する (S 0 8)。なお、解析データ D_1 および解析データ D_2 を、操作装置 1 3 に対する利用者からの指示に応じて編集してもよい。

40

【 0 0 2 6 】

図 2 の合成処理部 2 2 は、第 2 音信号 X_2 の解析データ D_2 を利用して第 1 音信号 X_1 の解析データ D_1 を変形する。本実施形態の合成処理部 2 2 は、アタック処理部 3 1 とリリース処理部 3 2 と音声合成部 3 3 とを含んで構成される。アタック処理部 3 1 は、第 2 音信号 X_2 におけるアタック部の音表現を第 1 音信号 X_1 に付加するアタック処理 S_1 を実行する。リリース処理部 3 2 は、第 2 音信号 X_2 におけるリリース部の音表現を第 1 音信号 X_1 に付加するリリース処理 S_2 を実行する。音声合成部 3 3 は、アタック処理部 3 1 およびリリース処理部 3 2 による処理後の解析データから変形音の第 3 音信号 Y を合成する。

50

【 0 0 2 7 】

図 5 には、歌唱音声の発音が開始された直後における基本周波数 f_1 の時間変化が図示されている。図 5 に例示される通り、定常期間 Q_1 の直前には有声期間 V_a が存在する。有声期間 V_a は、定常期間 Q_1 に先行する有声音の期間である。有声期間 V_a は、歌唱音声の音響特性（例えば基本周波数 f_1 またはスペクトル形状）が定常期間 Q_1 の直前に不安定に変動する期間である。例えば、歌唱音声の発音が開始した直後の定常期間 Q_1 に着目すると、歌唱音声の発音が開始される時刻 t_{1_A} から当該定常期間 Q_1 の始点時刻 T_{1_S} までのアタック部が有声期間 V_a に相当する。なお、以上の説明では歌唱音声に着目したが、参照音声についても同様に、定常期間 Q_2 の直前に有声期間 V_a が存在する。合成処理部 2 2（具体的にはアタック処理部 3 1）は、アタック処理 S_1 において、第 1 音信号 X_1 のうち有声期間 V_a と直後の定常期間 Q_1 とに対して第 2 音信号 X_2 におけるアタック部の音表現を付加する。

10

【 0 0 2 8 】

図 6 には、歌唱音声の発音が終了する直前における基本周波数 f_1 の時間変化が図示されている。図 6 に例示される通り、定常期間 Q_1 の直後には有声期間 V_r が存在する。有声期間 V_r は、定常期間 Q_1 に後続する有声音の期間である。有声期間 V_r は、歌唱音声の音響特性（例えば基本周波数 f_2 またはスペクトル形状）が定常期間 Q_1 の直後に不安定に変動する期間である。例えば、歌唱音声の発音が終了する直前の定常期間 Q_1 に着目すると、当該定常期間 Q_1 の終点時刻 T_{1_E} から歌唱音声が消音する時刻 t_{1_R} までのリリース部が有声期間 V_r に相当する。なお、以上の説明では歌唱音声に着目したが、参照音声についても同様に、定常期間 Q_2 の直後に音声期間 V_r が存在する。合成処理部 2 2（具体的にはリリース処理部 3 2）は、リリース処理 S_2 において、第 1 音信号 X_1 のうち有声期間 V_r と直前の定常期間 Q_1 とに対して第 2 音信号 X_2 のリリース部の音表現を付加する。

20

【 0 0 2 9 】

< リリース処理 S_2 >

図 7 は、リリース処理部 3 2 が実行するリリース処理 S_2 の具体的な内容を例示するフローチャートである。第 1 音信号 X_1 の定常期間 Q_1 毎に図 7 のリリース処理 S_2 が実行される。

【 0 0 3 0 】

リリース処理 S_2 を開始すると、リリース処理部 3 2 は、第 1 音信号 X_1 のうち処理対象の定常期間 Q_1 に第 2 音信号 X_2 のリリース部の音表現を付加するか否かを判定する（ S_{21} ）。具体的には、リリース処理部 3 2 は、以下に例示する条件 C_{r1} から条件 C_{r3} の何れかに該当する定常期間 Q_1 についてはリリース部の音表現を付加しないと判定する。ただし、第 1 音信号 X_1 の定常期間 Q_1 に音表現を付加するか否かを判定する条件は以下の例示に限定されない。

30

[条件 C_{r1}] 定常期間 Q_1 の時間長が所定値を下回る。

[条件 C_{r2}] 定常期間 Q_1 の直後の無声期間の時間長が所定値を下回る。

[条件 C_{r3}] 定常期間 Q_1 に後続する有声期間 V_r の時間長が所定値を上回る。

【 0 0 3 1 】

時間長が十分に短い定常期間 Q_1 には自然な声質で音表現を付加することが困難である。そこで、定常期間 Q_1 の時間長が所定値を下回る場合（条件 C_{r1} ）、リリース処理部 3 2 は、当該定常期間 Q_1 を音表現の付加対象から除外する。また、定常期間 Q_1 の直後に十分に短い無声期間が存在する場合、当該無声期間は、歌唱音声の途中における無声子音の期間である可能性がある。そして、無声子音の期間に音表現を付加すると、聴感的な違和感が知覚されるという傾向がある。以上の傾向を考慮して、定常期間 Q_1 の直後の無声期間の時間長が所定値を下回る場合（条件 C_{r2} ）、リリース処理部 3 2 は、当該定常期間 Q_1 を音表現の付加対象から除外する。また、定常期間 Q_1 の直後の有声期間 V_r の時間長が十分に長い場合には、歌唱音声に既に十分な音表現が付加されている可能性が高い。そこで、定常期間 Q_1 に後続する有声期間 V_r の時間長が十分に長い場合（条件 C_{r3} ）、リリース処理部 3 2 は、当該定常期間 Q_1 を音表現の付加対象から除外する。第 1 音信号 X_1 の定

40

50

常期間 Q1 に音表現を付加しないと判定した場合 (S21:NO)、リリース処理部 32 は、以下に詳述する処理 (S22 - S26) を実行することなくリリース処理 S2 を終了する。

【0032】

第 1 音信号 X1 の定常期間 Q1 に第 2 音信号 X2 のリリース部の音表現を付加すると判定した場合 (S21:YES)、リリース処理部 32 は、第 2 音信号 X2 の複数の定常期間 Q2 のうち、第 1 音信号 X1 の定常期間 Q1 に付加されるべき音表現に対応する定常期間 Q2 を選択する (S22)。具体的には、リリース処理部 32 は、処理対象の定常期間 Q1 に楽曲内の状況が近似する定常期間 Q2 を選択する。例えば、1 個の定常期間 (以下「着目定常期間」という) について考慮される状況 (context) としては、着目定常期間の時間長、着目定常期間の直後の定常期間の時間長、着目定常期間と直後の定常期間との間の音高差、着目定常期間の音高、および着目定常期間の直前の無音期間の時間長が例示される。リリース処理部 32 は、以上に例示した状況について定常期間 Q1 との差異が最小となる定常期間 Q2 を選択する。

10

【0033】

リリース処理部 32 は、以上の手順で選択した定常期間 Q2 に対応する音表現を第 1 音信号 X1 (解析データ D1) に付加するための処理 (S23 - S26) を実行する。図 8 は、リリース処理部 32 が第 1 音信号 X1 にリリース部の音表現を付加する処理の説明図である。

【0034】

図 8 には、第 1 音信号 X1 と第 2 音信号 X2 と変形後の第 3 音信号 Y との各々について、時間軸上の波形と基本周波数の時間変化とが併記されている。図 8 において、歌唱音声の定常期間 Q1 の始点時刻 T1_S および終点時刻 T1_E と、当該定常期間 Q1 の直後の有声期間 Vr の終点時刻 T1_R と、当該定常期間 Q1 の直後の音符に対応する有声期間 Va の始点時刻 T1_A と、参照音声の定常期間 Q2 の始点時刻 T2_S および終点時刻 T2_E と、当該定常期間 Q2 の直後の有声期間 Vr の終点時刻 T2_R とが、既知の情報である。

20

【0035】

リリース処理部 32 は、処理対象の定常期間 Q1 とステップ S22 で選択した定常期間 Q2 との間で時間軸上の位置関係を調整する (S23)。具体的には、リリース処理部 32 は、定常期間 Q2 の時間軸上の位置を、定常期間 Q1 の端点 (T1_S, T1_E) を基準とした位置に調整する。本実施形態のリリース処理部 32 は、図 8 に例示される通り、定常期間 Q1 の終点時刻 T1_E に定常期間 Q2 の終点時刻 T2_E が時間軸上で一致するように、第 2 音信号 X2 (定常期間 Q2) を第 1 音信号 X1 の時間軸上に配置する。

30

【0036】

< 処理期間 Z1_R の伸長 (S24) >

リリース処理部 32 は、第 1 音信号 X1 のうち第 2 音信号 X2 の音表現が付加される期間 (以下「処理期間」という) Z1_R を時間軸上で伸縮する (S24)。図 8 に例示される通り、処理期間 Z1_R は、音表現の付加が開始される時刻 (以下「合成開始時刻」という) Tm_R から定常期間 Q1 の直後の有声期間 Vr の終点時刻 T1_R までの期間である。合成開始時刻 Tm_R は、歌唱音声の定常期間 Q1 の始点時刻 T1_S と参照音声の定常期間 Q2 の始点時刻 T2_S とのうち後方の時刻である。図 8 の例示の通り、定常期間 Q2 の始点時刻 T2_S が定常期間 Q1 の始点時刻 T1_S の後方に位置する場合には、定常期間 Q2 の始点時刻 T2_S が合成開始時刻 Tm_R として設定される。ただし、合成開始時刻 Tm_R は始点時刻 T2_S に限定されない。

40

【0037】

図 8 に例示される通り、本実施形態のリリース処理部 32 は、第 1 音信号 X1 の処理期間 Z1_R を、第 2 音信号 X2 のうち表現期間 Z2_R の時間長に応じて伸長する。表現期間 Z2_R は、第 2 音信号 X2 のうちリリース部の音表現を表す期間であり、第 1 音信号 X1 に対する当該音表現の付加に利用される。図 8 に例示される通り、表現期間 Z2_R は、合成開始時刻 Tm_R から定常期間 Q2 の直後の有声期間 Vr の終点時刻 T2_R までの期間である。

【0038】

50

歌手等の熟練した歌唱者が歌唱した参照音声には相応の時間長にわたる十分な音表現が付加されるのに対し、歌唱に不慣れな利用者が歌唱した歌唱音声では音表現が時間的に不足する傾向がある。以上の傾向のもとでは、図 8 に例示される通り、参照音声の表現期間 Z_{2_R} が歌唱音声の処理期間 Z_{1_R} と比較して長い期間となる。したがって、本実施形態のリリース処理部 32 は、第 1 音信号 X_1 の処理期間 Z_{1_R} を、第 2 音信号 X_2 の表現期間 Z_{2_R} の時間長まで伸長する。

【0039】

処理期間 Z_{1_R} の伸長は、第 1 音信号 X_1 (歌唱音声) の任意の時刻 t_1 と変形後の第 3 音信号 Y (変形音) の任意の時刻 t とを相互に対応付ける処理 (マッピング) で実現される。図 8 には、歌唱音声の時刻 t_1 (縦軸) と変形音の時刻 t (横軸) との対応関係が図示されている。

10

【0040】

図 8 の対応関係における時刻 t_1 は、変形音の時刻 t に対応する第 1 音信号 X_1 の時刻である。図 8 に鎖線で併記された基準線 L は、第 1 音信号 X_1 が伸縮されない状態 ($t_1 = t$) を意味する。また、変形音の時刻 t に対する歌唱音声の時刻 t_1 の勾配が基準線 L と比較して小さい区間は、第 1 音信号 X_1 が伸長される区間を意味する。時刻 t に対する時刻 t_1 の勾配が基準線 L と比較して大きい区間は、歌唱音声が収縮される区間を意味する。

【0041】

時刻 t_1 と時刻 t との対応関係は、以下に例示する数式(1a)から数式(1c)の非線形関数で表現される。

20

【数 1】

$$t_1 = \begin{cases} t & (t < T_{_R}) \quad \dots (1a) \\ \eta \left(\frac{t - T_{_R}}{T_{1_R} - \tau_{2_R}} \right) (T_{1_R} - T_{_R}) + T_{_R} & (T_{_R} \leq t < \tau_{2_R}) \quad \dots (1b) \\ \frac{t - \tau_{2_R}}{T_{1_R} - \tau_{2_R}} (\tau_{1_A} - \tau_{1_R}) + \tau_{1_R} & (\tau_{2_R} \leq t < \tau_{1_A}) \quad \dots (1c) \end{cases}$$

30

【0042】

時刻 $T_{_R}$ は、図 8 に例示される通り、合成開始時刻 T_{m_R} と処理期間 Z_{1_R} の終点時刻 $1_{_R}$ との間に位置する所定の時刻である。例えば、定常期間 Q_1 の始点時刻 T_{1_S} と終点時刻 T_{1_E} との midpoint ($(T_{1_S} + T_{1_E}) / 2$) と合成開始時刻 T_{m_R} とのうちの後方の時刻が時刻 $T_{_R}$ として設定される。数式(1a)から理解される通り、処理期間 Z_{1_R} のうち時刻 $T_{_R}$ の前方の期間は伸縮されない。すなわち、時刻 $T_{_R}$ から処理期間 Z_{1_R} の伸長が開始される。

【0043】

数式(1b)から理解される通り、処理期間 Z_{1_R} のうち時刻 $T_{_R}$ の後方の期間は、当該時刻 $T_{_R}$ に近い位置において伸長の度合いが大きく、終点時刻 $1_{_R}$ に近づくほど伸長の度合いが小さくなるように時間軸上で伸長される。数式(1b)の関数 $\eta(t)$ は、時間軸上の前方ほど処理期間 Z_{1_R} を伸長し、時間軸上の後方ほど処理期間 Z_{1_R} の伸長の度合いを低減するための非線形関数である。具体的には、例えば時刻 t の 2 次関数 ($\eta(t) = t^2$) が関数 $\eta(t)$ として好適に利用される。以上に説明した通り、本実施形態では、処理期間 Z_{1_R} の終点時刻 $1_{_R}$ に近い位置ほど伸長の度合いが小さくなるように処理期間 Z_{1_R} が時間軸上で伸長される。したがって、歌唱音声の終点時刻 $1_{_R}$ の近傍の音響特性を変形音においても十分に維持することが可能である。なお、時刻 $T_{_R}$ に近い位置では、終点時刻 $1_{_R}$ の近傍と比較して、伸長に起因した聴感上の違和感が知覚され難い傾向がある。したがって、前述の例示のように時刻 $T_{_R}$ に近い位置において伸長の度合いを増大させても、変形音の聴感上の自然性は殆ど低下しない。なお、第 1 音信号 X_1 のうち表現期間 Z_{2_R} の終点時刻

40

50

2_Rから次の有声期間V_rの始点時刻 1_Aまでの期間は数式(1c)から理解される通り時間軸上で短縮される。なお、終点時刻 2_Rから始点時刻 1_Aまでの期間には音声が存在しないから、第1音信号X₁を部分的な削除により削除してもよい。

【0044】

以上の例示の通り、歌唱音声の処理期間Z_{1_R}は参照音声の表現期間Z_{2_R}の時間長に伸長される。他方、参照音声の表現期間Z_{2_R}は時間軸上で伸縮されない。すなわち、変形音の時刻tに対応する配置後の第2音信号X₂の時刻t₂は当該時刻tに一致する(t₂ = t)。以上の例示の通り、本実施形態においては、歌唱音声の処理期間Z_{1_R}が表現期間Z_{2_R}の時間長に応じて伸長されるから、第2音信号X₂の伸長は不要である。したがって、第2音信号X₂が表すリリース部の音表現を正確に第1音信号X₁に付加することが可能である。

10

【0045】

以上に例示した手順で処理期間Z_{1_R}を伸長すると、リリース処理部32は、第1音信号X₁の伸長後の処理期間Z_{1_R}を第2音信号X₂の表現期間Z_{2_R}に応じて変形する(S25 - S26)。具体的には、歌唱音声の伸長後の処理期間Z_{1_R}と参照音声の表現期間Z_{2_R}との間で、基本周波数の合成(S25)とスペクトル包絡概形の合成(S26)とが実行される。

【0046】

<基本周波数の合成(S25)>

リリース処理部32は、以下の数式(2)の演算により第3音信号Yの各時刻tにおける基本周波数F(t)を算定する。

20

【数2】

$$F(t) = f_1(t_1) - \lambda_1(f_1(t_1) - F_1(t_1)) + \lambda_2(f_2(t_2) - F_2(t_2)) \quad \dots (2)$$

【0047】

数式(2)における平滑基本周波数F₁(t₁)は、第1音信号X₁の基本周波数f₁(t₁)の時系列を時間軸上で平滑化した周波数である。同様に、数式(2)の平滑基本周波数F₂(t₂)は、第2音信号X₂の基本周波数f₂(t₂)の時系列を時間軸上で平滑化した周波数である。数式(2)の係数₁および係数₂は1以下の非負値に設定される(0 ≤ ₁ ≤ 1, 0 ≤ ₂ ≤ 1)。

30

【0048】

数式(2)から理解される通り、数式(2)の第2項は、歌唱音声の基本周波数f₁(t₁)と平滑基本周波数F₁(t₁)との差分を、係数₁に応じた度合で、第1音信号X₁の基本周波数f₁(t₁)から低減する処理である。また、数式(2)の第3項は、参照音声の基本周波数f₂(t₂)と平滑基本周波数F₂(t₂)との差分を、係数₂に応じた度合で、第1音信号X₁の基本周波数f₁(t₁)に付加する処理である。以上の説明から理解される通り、リリース処理部32は、歌唱音声の基本周波数f₁(t₁)と平滑基本周波数F₁(t₁)との差分を、参照音声の基本周波数f₂(t₂)と平滑基本周波数F₂(t₂)との差分に置換する要素として機能する。すなわち、第1音信号X₁における伸長後の処理期間Z_{1_R}内の基本周波数f₁(t₁)の時間変化が、第2音信号X₂における表現期間Z_{2_R}内の基本周波数f₂(t₂)の時間変化に近付く。

40

【0049】

<スペクトル包絡概形の合成(S26)>

リリース処理部32は、歌唱音声の伸長後の処理期間Z_{1_R}と参照音声の表現期間Z_{2_R}との間でスペクトル包絡概形を合成する。第1音信号X₁のスペクトル包絡概形G₁は、図9に例示される通り、第1音信号X₁の周波数スペクトルg₁の概形であるスペクトル包絡g₂を周波数領域で更に平滑化した強度分布を意味する。具体的には、音韻性(音韻に依存した差異)および個人性(発声者に依存した差異)が知覚できなくなる程度にスペクトル包絡g₂を平滑化した強度分布がスペクトル包絡概形G₁である。例えばスペクトル包

50

絡 g_2 を表すメルケプストラムの複数の係数のうち低次側に位置する所定個の係数によりスペクトル包絡概形 G_1 が表現される。以上の説明では第 1 音信号 X_1 のスペクトル包絡概形 G_1 に着目したが、第 2 音信号 X_2 のスペクトル包絡概形 G_2 も同様である。

【 0 0 5 0 】

リリース処理部 3 2 は、以下の数式(3)の演算により第 3 音信号 Y の各時刻 t におけるスペクトル包絡概形 (以下「合成スペクトル包絡概形」という) $G(t)$ を算定する。

【数 3】

$$G(t) = G_1(t_1) - \mu_1(G_1(t_1) - G_{1_ref}) + \mu_2(G_2(t_2) - G_{2_ref}) \dots (3)$$

10

【 0 0 5 1 】

数式(3)の記号 G_{1_ref} は、基準スペクトル包絡概形である。第 1 音信号 X_1 の複数のスペクトル包絡概形 G_1 のうち、特定の時点における 1 個のスペクトル包絡概形 G_1 が、基準スペクトル包絡概形 G_{1_ref} (第 1 基準スペクトル包絡概形の例示) として利用される。具体的には、基準スペクトル包絡概形 G_{1_ref} は、第 1 音信号 X_1 のうち合成開始時刻 T_{m_R} (第 1 時点の例示) におけるスペクトル包絡概形 $G_1(T_{m_R})$ である。すなわち、基準スペクトル包絡概形 G_{1_ref} が抽出される時点は、定常期間 Q_1 の始点時刻 T_{1_S} および定常期間 Q_2 の始点時刻 T_{2_S} のうち後方の時刻に位置する。なお、基準スペクトル包絡概形 G_{1_ref} が抽出される時点は合成開始時刻 T_{m_R} に限定されない。例えば、定常期間 Q_1 内の任意の時点のスペクトル包絡概形 G_1 が基準スペクトル包絡概形 G_{1_ref} として利用される。

20

【 0 0 5 2 】

同様に、数式(3)の基準スペクトル包絡概形 G_{2_ref} は、第 2 音信号 X_2 の複数のスペクトル包絡概形 G_2 のうち、特定の時点における 1 個のスペクトル包絡概形 G_2 である。具体的には、基準スペクトル包絡概形 G_{2_ref} は、第 2 音信号 X_2 のうち合成開始時刻 T_{m_R} (第 2 時点の例示) におけるスペクトル包絡概形 $G_2(T_{m_R})$ である。すなわち、基準スペクトル包絡概形 G_{2_ref} が抽出される時点は、定常期間 Q_1 の始点時刻 T_{1_S} および定常期間 Q_2 の始点時刻 T_{2_S} のうち後方の時刻に位置する。なお、基準スペクトル包絡概形 G_{2_ref} が抽出される時点は合成開始時刻 T_{m_R} に限定されない。例えば、定常期間 Q_1 内の任意の時点のスペクトル包絡概形 G_2 が基準スペクトル包絡概形 G_{2_ref} として利用される。

30

【 0 0 5 3 】

数式(3)の係数 μ_1 および係数 μ_2 は、1 以下の非負値に設定される ($0 \leq \mu_1 \leq 1, 0 \leq \mu_2 \leq 1$)。数式(3)の第 2 項は、歌唱音声のスペクトル包絡概形 $G_1(t_1)$ と基準スペクトル包絡概形 G_{1_ref} との差分を、係数 μ_1 (第 1 係数の例示) に応じた度合で、第 1 音信号 X_1 のスペクトル包絡概形 $G_1(t_1)$ から低減する処理である。また、数式(3)の第 3 項は、参照音声のスペクトル包絡概形 $G_2(t_2)$ と基準スペクトル包絡概形 G_{2_ref} との差分を、係数 μ_2 (第 2 係数の例示) に応じた度合で、第 2 音信号 X_2 のスペクトル包絡概形 $G_2(b)$ から低減する処理である。以上の説明から理解される通り、リリース処理部 3 2 は、歌唱音声のスペクトル包絡概形 $G_1(t_1)$ と基準スペクトル包絡概形 G_{1_ref} との差分 (第 1 差分の例示) を、参照音声のスペクトル包絡概形 $G_2(t_2)$ と基準スペクトル包絡概形 G_{2_ref} との差分 (第 2 差分の例示) に置換する要素として機能する。

40

【 0 0 5 4 】

< アタック処理 S 1 >

図 1 0 は、アタック処理部 3 1 が実行するアタック処理 S 1 の具体的な内容を例示するフローチャートである。第 1 音信号 X_1 の定常期間 Q_1 毎に図 1 0 のアタック処理 S 1 が実行される。なお、アタック処理 S 1 の具体的な手順はリリース処理 S 2 と同様である。

【 0 0 5 5 】

アタック処理 S 1 を開始すると、アタック処理部 3 1 は、第 1 音信号 X_1 のうち処理対象の定常期間 Q_1 に第 2 音信号 X_2 のアタック部の音表現を付加するか否かを判定する (S 1

50

1)。具体的には、アタック処理部 3 1 は、以下に例示する条件 C a 1 から条件 C a 5 の何れかに該当する定常期間 Q 1 についてはアタック部の音表現を付加しないと判定する。ただし、第 1 音信号 X 1 の定常期間 Q 1 に音表現を付加するか否かを判定する条件は以下の例示に限定されない。

[条件 C a 1] 定常期間 Q 1 の時間長が所定値を下回る。

[条件 C a 2] 定常期間 Q 1 内で平滑化した基本周波数 f 1 の変動幅が所定値を上回る。

[条件 C a 3] 定常期間 Q 1 のうち始点を含む所定長の期間内で平滑化した基本周波数 f 1 の変動幅が所定値を上回る。

[条件 C a 4] 定常期間 Q 1 の直前の有声期間 V a の時間長が所定値を上回る。

[条件 C a 5] 定常期間 Q 1 の直前の有声期間 V a における基本周波数 f 1 の変動幅が所定値を上回る。

10

【0056】

条件 C a 1 は、前述の条件 C r 1 と同様に、時間長が十分に短い定常期間 Q 1 には自然な声質で音表現を付加することが困難であるという事情を考慮した条件である。また、定常期間 Q 1 内で基本周波数 f 1 が大きく変動する場合には、歌唱音声に十分な音表現が付加されている可能性が高い。そこで、平滑後の基本周波数 f 1 の変動幅が所定値を上回る定常期間 Q 1 は、音表現の付加対象から除外される(条件 C a 2)。条件 C a 3 は、条件 C a 2 と同様の内容であるが、定常期間 Q 1 のうち特にアタック部に近い期間に着目した条件である。

また、定常期間 Q 1 の直前の有声期間 V a の時間長が十分に長い場合、または有声期間 V a 内で基本周波数 f 1 が大きく変動する場合には、歌唱音声に既に十分な音表現が付加されている可能性が高い。そこで、直前の有声期間 V a の時間長が所定値を上回る定常期間 Q 1 (条件 C a 4) と、有声期間 V a 内での基本周波数 f 1 の変動幅が所定値を上回る定常期間 Q 1 (条件 C a 5) とは、音表現の付加対象から除外される。定常期間 Q 1 に音表現を付加しないと判定した場合(S 11: YES)、アタック処理部 3 1 は、以下に詳述する処理(S 12 - S 16)を実行することなくアタック処理 S 1 を終了する。

20

【0057】

第 1 音信号 X 1 の定常期間 Q 1 に第 2 音信号 X 2 のアタック部の音表現を付加すると判定した場合(S 11: YES)、アタック処理部 3 1 は、第 2 音信号 X 2 の複数の定常期間 Q 2 のうち、定常期間 Q 1 に付加されるべき音表現に対応する定常期間 Q 2 を選択する(S 12)。アタック処理部 3 1 が定常期間 Q 2 を選択する方法は、リリース処理部 3 2 が定常期間 Q 2 を選択する方法と同様である。

30

【0058】

アタック処理部 3 1 は、以上の手順で選択した定常期間 Q 2 に対応する音表現を第 1 音信号 X 1 に付加するための処理(S 13 - S 16)を実行する。図 1 1 は、アタック処理部 3 1 が第 1 音信号 X 1 にアタック部の音表現を付加する処理の説明図である。

【0059】

アタック処理部 3 1 は、処理対象の定常期間 Q 1 とステップ S 12 で選択した定常期間 Q 2 との間で時間軸上の位置関係を調整する(S 13)。具体的には、アタック処理部 3 1 は、図 1 1 に例示される通り、定常期間 Q 1 の始点時刻 T 1_S に定常期間 Q 2 の始点時刻 T 2_S が時間軸上で一致するように、第 2 音信号 X 2 (定常期間 Q 2) を第 1 音信号 X 1 の時間軸上に配置する。

40

【0060】

<処理期間 Z 1_A の伸長>

アタック処理部 3 1 は、第 1 音信号 X 1 のうち第 2 音信号 X 2 の音表現が付加される処理期間 Z 1_A を時間軸上で伸長する(S 14)。処理期間 Z 1_A は、定常期間 Q 1 の直前の有声期間 V a の始点時刻 T 1_A から音表現の付加が終了される時刻(以下「合成終了時刻」という) T m_A までの期間である。合成終了時刻 T m_A は、例えば定常期間 Q 1 の始点時刻 T 1_S (定常期間 Q 2 の始点時刻 T 2_S) である。すなわち、アタック処理 S 1 においては、定常期間 Q 1 の前方の有声期間 V a が処理期間 Z 1_A として伸長される。前述の通り、定常期間 Q 1 は楽曲の音符に相当する期間である。有声期間 V a を伸長し、定常期間 Q 1 は伸

50

長しない構成によれば、定常期間 Q1 の始点時刻 $T1_S$ の変化が抑制される。すなわち、歌唱音声における音符の先頭が前後に移動する可能性を低減できる。

【0061】

図 11 に例示される通り、本実施形態のアタック処理部 31 は、第 1 音信号 X1 の処理期間 $Z1_A$ を、第 2 音信号 X2 のうち表現期間 $Z2_A$ の時間長に応じて伸長する。表現期間 $Z2_A$ は、第 2 音信号 X2 のうちアタック部の音表現を表す期間であり、第 1 音信号 X1 に対する当該音表現の付加に利用される。図 11 に例示される通り、表現期間 $Z2_A$ は、定常期間 Q2 の直前の有声期間 Va である。

【0062】

具体的には、アタック処理部 31 は、第 1 音信号 X1 の処理期間 $Z1_A$ を、第 2 音信号 X2 の表現期間 $Z2_A$ の時間長まで伸長する。図 11 には、歌唱音声の時刻 $t1$ (縦軸) と変形音の時刻 t (横軸) との対応関係が図示されている。

10

【0063】

図 11 に例示される通り、本実施形態では、処理期間 $Z1_A$ の始点時刻 $t1_A$ に近い位置ほど伸長の度合が小さくなるように処理期間 $Z1_A$ が時間軸上で伸長される。したがって、歌唱音声の始点時刻 $t1_A$ の近傍の音響特性を変形音においても十分に維持することが可能である。他方、参照音声の表現期間 $Z2_A$ は時間軸上で伸縮されない。したがって、第 2 音信号 X2 が表すアタック部の音表現を正確に第 1 音信号 X1 に付加することが可能である。

【0064】

20

以上に例示した手順で処理期間 $Z1_A$ を伸長すると、アタック処理部 31 は、第 1 音信号 X1 の伸長後の処理期間 $Z1_A$ を第 2 音信号 X2 の表現期間 $Z2_A$ に応じて変形する (S15 - S16)。具体的には、歌唱音声の伸長後の処理期間 $Z1_A$ と参照音声の表現期間 $Z2_A$ との間で、基本周波数の合成 (S25) とスペクトル包絡概形の合成 (S26) とが実行される。

【0065】

具体的には、アタック処理部 31 は、前述の数式(2)と同様の演算により、第 1 音信号 X1 の基本周波数 $f1(t1)$ と第 2 音信号 X2 の基本周波数 $f2(t2)$ とから第 3 音信号 Y の基本周波数 $F(t)$ を算定する。すなわち、アタック処理部 31 は、基本周波数 $f1(t1)$ と平滑後の基本周波数 $F1(t1)$ との差分を係数 1 に応じた度合で第 1 音信号 X1 の基本周波数 $f1(t1)$ から低減し、基本周波数 $f2(t2)$ と平滑後の基本周波数 $F2(t2)$ との差分を係数 2 に応じた度合で第 1 音信号 X1 の基本周波数 $f1(t1)$ に付加することで、第 3 音信号 Y の基本周波数 $F(t)$ を算定する。したがって、第 1 音信号 X1 における伸長後の処理期間 $Z1_A$ 内の基本周波数 $f1(t1)$ の時間変化が、第 2 音信号 X2 における表現期間 $Z2_A$ 内の基本周波数 $f2(t2)$ の時間変化に近づく。

30

【0066】

また、アタック処理部 31 は、歌唱音声の伸長後の処理期間 $Z1_A$ と参照音声の表現期間 $Z2_A$ との間でスペクトル包絡概形を合成する。具体的には、アタック処理部 31 は、前述の数式(3)と同様の演算により、第 1 音信号 X1 のスペクトル包絡概形 $G1(t1)$ と第 2 音信号 X2 のスペクトル包絡概形 $G2(t2)$ とから第 3 音信号 Y の合成スペクトル包絡概形 $G(t)$ を算定する。アタック処理 S1 において数式(3)に適用される基準スペクトル包絡概形 $G1_ref$ は、第 1 音信号 X1 のうち合成終了時刻 Tm_A (第 1 時点の例示) におけるスペクトル包絡概形 $G1(Tm_A)$ である。すなわち、基準スペクトル包絡概形 $G1_ref$ が抽出される時点は、定常期間 Q1 の始点時刻 $T1_S$ に位置する。

40

【0067】

同様に、アタック処理 S1 において数式(3)に適用される基準スペクトル包絡概形 $G2_ref$ は、第 2 音信号 X2 のうち合成終了時刻 Tm_A (第 2 時点の例示) におけるスペクトル包絡概形 $G2(Tm_A)$ である。すなわち、基準スペクトル包絡概形 $G2_ref$ が抽出される時点は、定常期間 Q1 の始点時刻 $T1_S$ に位置する。

【0068】

50

以上の説明から理解される通り、本実施形態のアタック処理部 3 1 およびリリース処理部 3 2 の各々は、定常期間 Q 1 の端点（始点時刻 T 1_S または終点時刻 T 1_E）を基準とした時間軸上の位置において第 2 音信号 X 2（解析データ D 2）を利用して第 1 音信号 X 1（解析データ D 1）を変形する。以上に例示したアタック処理 S 1 およびリリース処理 S 2 により、変形音を表す第 3 音信号 Y の基本周波数 F (t) の時系列と合成スペクトル包絡概形 G (t) の時系列とが生成される。図 2 の音声合成部 3 3 は、第 3 音信号 Y の基本周波数 F (t) の時系列と合成スペクトル包絡概形 G (t) の時系列とから第 3 音信号 Y を生成する。

【 0 0 6 9 】

図 2 の音声合成部 3 3 は、アタック処理 S 1 およびリリース処理 S 2 の結果（すなわち変形後の解析データ）を利用して変形音の第 3 音信号 Y を合成する。具体的には、音声合成部 3 3 は、第 1 音信号 X 1 から算定される各周波数スペクトル g 1 を合成スペクトル包絡概形 G (t) に沿うように調整し、かつ、第 1 音信号 X 1 の基本周波数 f 1 を基本周波数 F (t) に調整する。周波数スペクトル g 1 および基本周波数 f 1 の調整は例えば周波数領域で実行される。音声合成部 3 3 は、以上に例示した調整後の周波数スペクトルを時間領域に変換することで第 3 音信号 Y を合成する。

10

【 0 0 7 0 】

以上に説明した通り、本実施形態では、第 1 音信号 X 1 のスペクトル包絡概形 G 1 (t 1) と基準スペクトル包絡概形 G 1_ref との差分 (G 1 (t 1) - G 1_ref) と、第 2 音信号 X 2 のスペクトル包絡概形 G 2 (t 2) と基準スペクトル包絡概形 G 2_ref との差分 (G 2 (t 2) - G 2_ref) とが、第 1 音信号 X 1 のスペクトル包絡概形 G 1 (t 1) に合成される。したがって、第 1 音信号 X 1 のうち、第 2 音信号 X 2 を利用して変形される期間（処理期間 Z 1_A , Z 1_R）と当該期間の前後の期間との境界において音響特性が連続する聴感的に自然な変形音を生成できる。

20

【 0 0 7 1 】

また、本実施形態では、第 1 音信号 X 1 のうち基本周波数 f 1 およびスペクトル形状が時間的に安定している定常期間 Q 1 が特定され、定常期間 Q 1 の端点（始点時刻 T 1_S または終点時刻 T 1_E）を基準として配置された第 2 音信号 X 2 を利用して第 1 音信号 X 1 が変形される。したがって、第 1 音信号 X 1 の適切な期間が第 2 音信号 X 2 に応じて変形され、聴感的に自然な変形音を生成できる。

【 0 0 7 2 】

本実施形態では、第 1 音信号 X 1 の処理期間（ Z 1_A , Z 1_R ）が第 2 音信号 X 2 の表現期間（ Z 2_A , Z 2_R ）の時間長に応じて伸長されるから、第 2 音信号 X 2 の伸長は不要である。したがって、参照音声の音響特性（例えば音表現）が正確に第 1 音信号 X 1 に付加され、聴感的に自然な変形音を生成できる。

30

【 0 0 7 3 】

< 変形例 >

以上に例示した各態様に付加される具体的な変形の態様を以下に例示する。以下の例示から任意に選択された 2 個以上の態様を、相互に矛盾しない範囲で適宜に併合してもよい。

【 0 0 7 4 】

(1) 前述の形態では、第 1 指標 1 と第 2 指標 2 とから算定される変動指標 を利用して第 1 音信号 X 1 の定常期間 Q 1 を特定したが、第 1 指標 1 と第 2 指標 2 とに応じて定常期間 Q 1 を特定する方法は以上の例示に限定されない。例えば、信号解析部 2 1 は、第 1 指標 1 に応じた第 1 暫定期間と第 2 指標 2 に応じた第 2 暫定期間とを特定する。第 1 暫定期間は、例えば第 1 指標 1 が閾値を下回る有声音の期間である。すなわち、基本周波数 f 1 が時間的に安定している期間が第 1 暫定期間として特定される。第 2 暫定期間は、例えば第 2 指標 2 が閾値を下回る有声音の期間である。すなわち、スペクトル形状が時間的に安定している期間が第 2 暫定期間として特定される。信号解析部 2 1 は、第 1 暫定期間と第 2 暫定期間とが相互に重複する期間を定常期間 Q 1 として特定する。すなわち、第 1 音信号 X 1 のうち基本周波数 f 1 とスペクトル形状との双方が時間的に安定している期間が定常期間 Q 1 として特定される。以上の説明から理解される通り、定常期間 Q 1 の特

40

50

定において変動指標 の算定を省略してもよい。なお、以上の説明では定常期間 Q1 の特定に着目したが、第 2 音信号 X2 における定常期間 Q2 の特定についても同様である。

【 0 0 7 5 】

(2) 前述の形態では、第 1 音信号 X1 のうち基本周波数 f_1 およびスペクトル形状の双方が時間的に安定する期間を定常期間 Q1 として特定したが、第 1 音信号 X1 のうち基本周波数 f_1 およびスペクトル形状の一方が時間的に安定する期間を定常期間 Q1 として特定してもよい。同様に、第 2 音信号 X2 のうち基本周波数 f_2 およびスペクトル形状の一方が時間的に安定する期間を定常期間 Q2 として特定してもよい。

【 0 0 7 6 】

(3) 前述の形態では、第 1 音信号 X1 のうち合成開始時刻 T_{m_R} または合成終了時刻 T_{m_A} におけるスペクトル包絡概形 G1 を基準スペクトル包絡概形 G1_ref として利用したが、基準スペクトル包絡概形 G1_ref が抽出される時点 (第 1 時点) は以上の例示に限定されない。例えば、定常期間 Q1 の端点 (始点時刻 T_{1_S} または終点時刻 T_{1_E}) におけるスペクトル包絡概形 G1 を基準スペクトル包絡概形 G1_ref としてもよい。ただし、基準スペクトル包絡概形 G1_ref が抽出される第 1 時点は、第 1 音信号 X1 のうちスペクトル形状が安定している定常期間 Q1 内の時点であることが望ましい。

【 0 0 7 7 】

基準スペクトル包絡概形 G2_ref についても同様である。すなわち、前述の形態では、第 2 音信号 X2 のうち合成開始時刻 T_{m_R} または合成終了時刻 T_{m_A} におけるスペクトル包絡概形 G2 を基準スペクトル包絡概形 G2_ref として利用したが、基準スペクトル包絡概形 G2_ref が抽出される時点 (第 2 時点) は以上の例示に限定されない。例えば、定常期間 Q2 の端点 (始点時刻 T_{2_S} または終点時刻 T_{2_E}) におけるスペクトル包絡概形 G2 を基準スペクトル包絡概形 G2_ref としてもよい。ただし、基準スペクトル包絡概形 G2_ref が抽出される第 2 時点は、第 2 音信号 X2 のうちスペクトル形状が安定している定常期間 Q2 内の時点であることが望ましい。

【 0 0 7 8 】

また、第 1 音信号 X1 のうち基準スペクトル包絡概形 G1_ref が抽出される第 1 時点と、第 2 音信号 X2 のうち基準スペクトル包絡概形 G2_ref が抽出される第 2 時点とは、時間軸上の相異なる時点でもよい。

【 0 0 7 9 】

(4) 前述の形態では、音処理装置 100 の利用者が歌唱した歌唱音声を表す第 1 音信号 X1 を処理したが、第 1 音信号 X1 が表す音声は、利用者による歌唱音声に限定されない。例えば、素片接続型または統計モデル型の公知の音声合成技術により合成された第 1 音信号 X1 を処理してもよい。また、光ディスク等の記録媒体から読出された第 1 音信号 X1 を処理してもよい。第 2 音信号 X2 についても同様に、任意の方法で取得される。

【 0 0 8 0 】

また、第 1 音信号 X1 および第 2 音信号 X2 が表す音響は、狭義の音声 (すなわち人間が発声する言語音) に限定されない。例えば、楽器の演奏音を表す第 1 音信号 X1 に各種の音表現 (例えば演奏表現) を付加する場合にも本発明は適用される。例えば、演奏表現が付加されていない単調な演奏音を表す第 1 音信号 X1 に対し、第 2 音信号 X2 を利用してピブラート等の演奏表現が付加される。

【 0 0 8 1 】

< 付記 >

以上に例示した形態から、例えば以下の構成が把握される。

【 0 0 8 2 】

本発明の好適な態様 (第 1 態様) に係る音声処理方法は、歌唱音声を表す第 1 音信号のうちの処理期間を、前記歌唱音声とは音響特性が相違する参照音声を表す前記第 2 音信号において前記第 1 音信号の変形に適用されるべき表現期間の時間長に応じて伸長し、前記処理期間の伸長後の前記第 1 音信号を、前記第 2 音信号の前記表現期間に応じて変形合成する。以上の態様では、第 1 音信号のうちの処理期間が第 2 音信号の表現期間の時間長に

10

20

30

40

50

応じて伸長されるから、参照音声を表す第2音信号の伸縮は不要である。したがって、参照音声の音響特性を正確に第1音信号に付加することが可能である。

【0083】

第1態様の好適例(第2態様)において、前記処理期間は、前記歌唱音声のリリース部を含む期間であり、前記処理期間の伸長においては、当該処理期間の終点に近い位置ほど伸長の度合いが小さくなるように前記処理期間を伸長する。以上の態様によれば、処理期間の終点に近い位置ほど伸長の度合いが小さくなるから、歌唱音声の終点の近傍の音響特性を維持しながら参照音声の音響特性を付加することが可能である。

【0084】

第1態様の好適例(第3態様)において、前記処理期間は、前記歌唱音声のアタック部を含む期間であり、前記処理期間の伸長においては、前記処理期間の始点に近い位置ほど伸長の度合いが小さくなるように前記処理期間を伸長する。以上の態様によれば、処理期間の始点に近い位置ほど伸長の度合いが小さくなるから、歌唱音声の始点の近傍の音響特性を維持しながら参照音声の音響特性を付加することが可能である。

10

【0085】

本発明の好適な態様(第4態様)に係る音声処理装置は、歌唱音声を表す第1音信号のうちの処理期間を、前記歌唱音声とは音響特性が相違する参照音声を表す前記第2音信号において前記第1音信号の変形に適用されるべき表現期間の時間長に応じて伸長し、前記処理期間の伸長後の前記第1音信号を、前記第2音信号の前記表現期間に応じて変形する合成処理部を具備する。

20

【0086】

第4態様の好適例(第5態様)において、前記処理期間は、前記歌唱音声のリリース部を含む期間であり、前記合成処理部は、当該処理期間の終点に近い位置ほど伸長の度合いが小さくなるように前記処理期間を伸長する。

【0087】

第4態様の好適例(第6態様)において、前記処理期間は、前記歌唱音声のアタック部を含む期間であり、前記合成処理部は、前記処理期間の始点に近い位置ほど伸長の度合いが小さくなるように前記処理期間を伸長する。

【符号の説明】

【0088】

100...音声処理装置、11...制御装置、12...記憶装置、13...操作装置、14...放音装置、21...信号解析部、22...合成処理部、31...アタック処理部、32...リリース処理部、33...音声合成部。

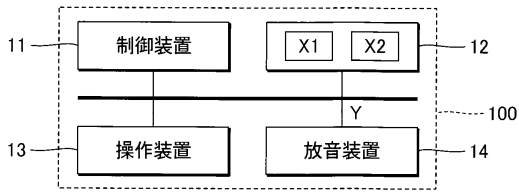
30

40

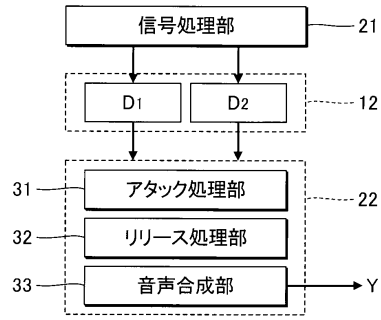
50

【図面】

【図 1】

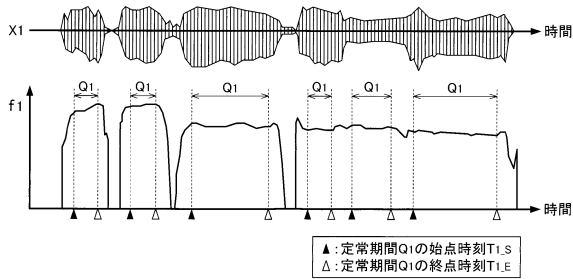


【図 2】

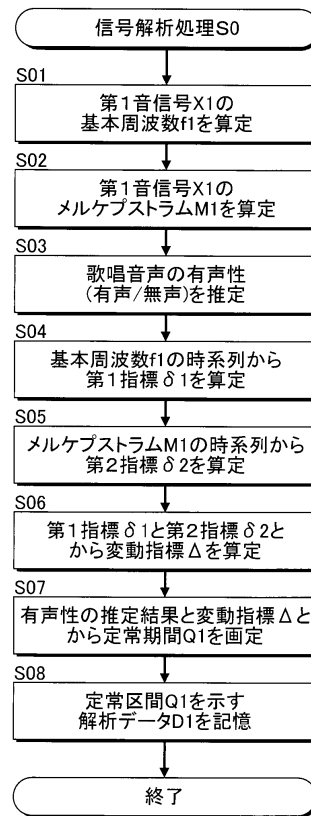


10

【図 3】



【図 4】



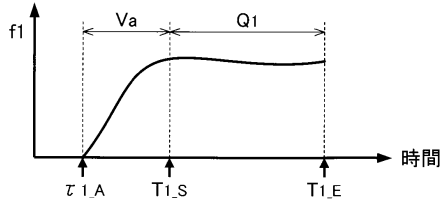
20

30

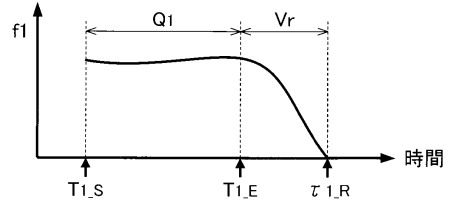
40

50

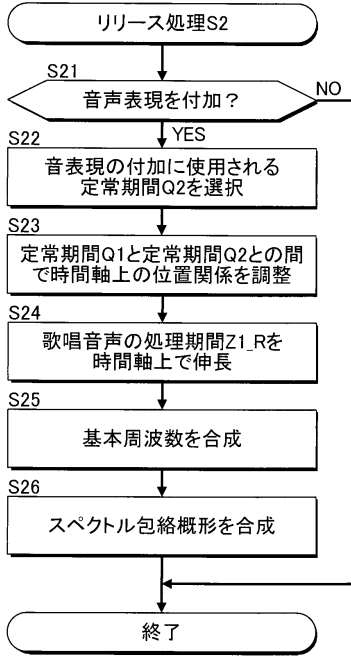
【図5】



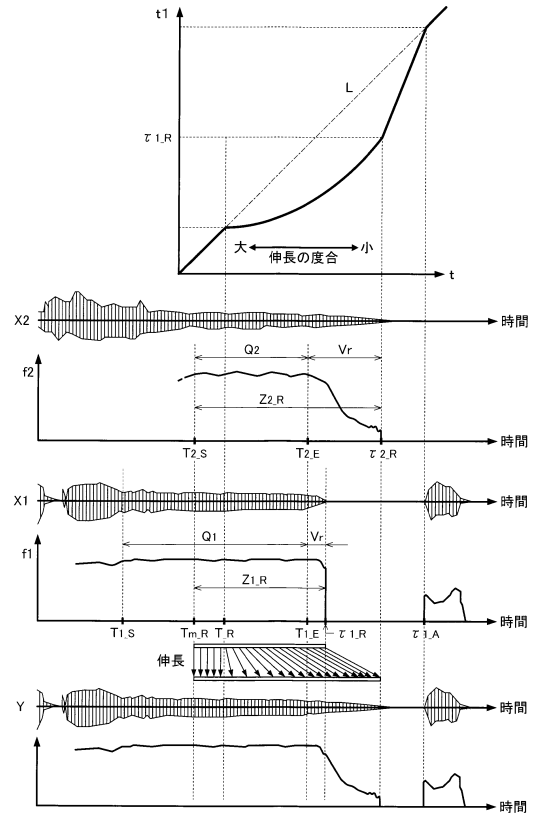
【図6】



【図7】



【図8】



10

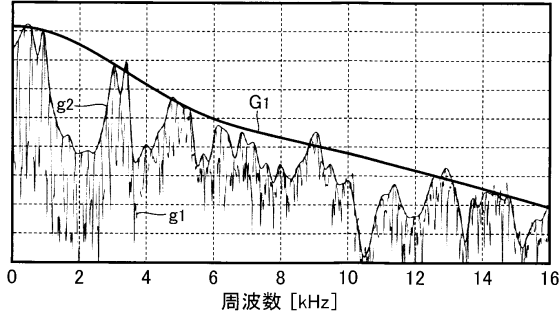
20

30

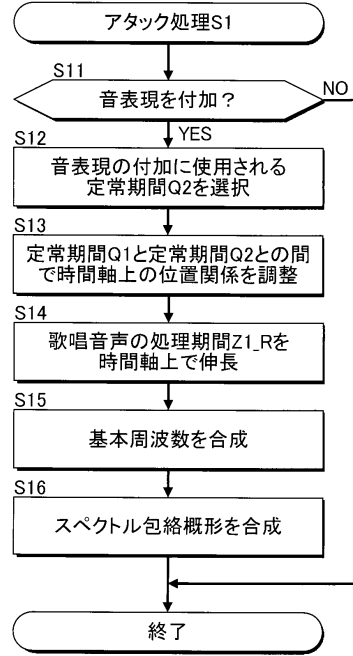
40

50

【図9】



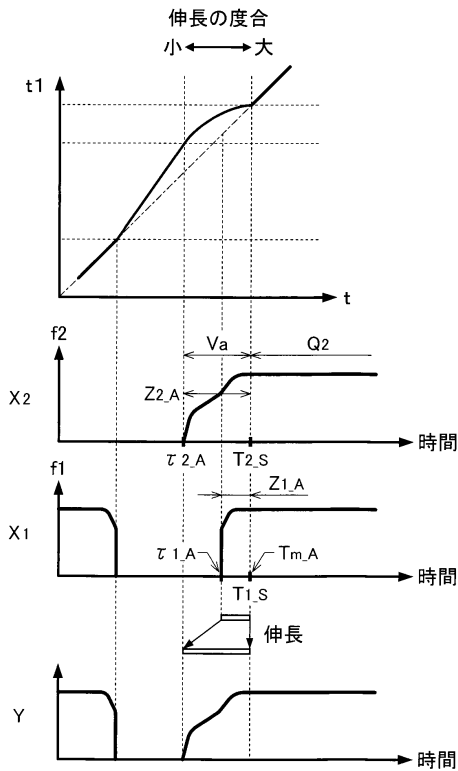
【図10】



10

20

【図11】



30

40

50

フロントページの続き

審査官 菊池 智紀

- (56)参考文献 特開2001-242900(JP,A)
国際公開第2008/142836(WO,A1)
- (58)調査した分野 (Int.Cl., DB名)
G10L 13/00 - 13/10,
21/003 - 21/013,
21/04 - 21/057
G10H 1/00 - 7/12
G10K 15/04