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

特許第5360489号
(P5360489)

(45)発行日　平成25年12月4日(2013.12.4)　(24)登録日　平成25年9月13日(2013.9.13)

(51)Int.Cl.　F I

G10L 13/06	(2013.01)	G10L	13/06	140
G10L 13/033	(2013.01)	G10L	13/02	122B
G10L 19/00	(2013.01)	G10L	19/00	220Z

請求項の数 15　(全 22 頁)

(21)出願番号　特願2009-244698(P2009-244698)
(22)出願日　平成21年10月23日(2009.10.23)
(65)公開番号　特開2011-90218(P2011-90218A)
(43)公開日　平成23年5月6日(2011.5.6)
審査請求日　平成24年8月8日(2012.8.8)

(73)特許権者　000002897
大日本印刷株式会社
東京都新宿区市谷加賀町一丁目１番１号
(74)代理人　100111659
弁理士　金山　聡
(74)代理人　100135954
弁理士　深町　圭子
(74)代理人　100119057
弁理士　伊藤　英生
(74)代理人　100122529
弁理士　藤枡　裕実
(74)代理人　100131369
弁理士　後藤　直樹

最終頁に続く

(54)【発明の名称】音素符号変換装置および音声合成装置

(57)【特許請求の範囲】
【請求項１】
発声された１つの音節を所定のサンプリング周波数でサンプル数Ｊの音節波形信号として取得する音節波形取得手段と、
前記音節波形信号を当該サンプリング周波数を維持しながら時間軸上でＱ（Ｑは２以上の整数）倍拡大し、Ｊ×Ｑサンプルの拡大波形信号に変換する音節波形拡大手段と、
前記拡大波形信号に対して所定の周波数解析を行い、発音開始時刻、発音終了時刻、音高、強さのデータを備える複数の符号コードで表現した符号コード群を作成する符号コード群作成手段と、
前記符号コード群を構成する全ての符号コードの音高をＱ倍に補正し、全ての符号コードの発音開始時刻と発音終了時刻を１／Ｑ倍に補正し、補正された符号コード群で構成される音節符号を作成する音節符号作成手段と、
前記音節符号を構成する各符号コードについて、各音高ごとに発音開始時刻と発音終了時刻との時間差と符号コードの強さとの積で与えられるエネルギー値の総和であるエネルギー総和値を算出する音高別エネルギー算出手段と、
複数の音節符号間で、各音高ごとに全てのエネルギー総和値を乗算して統合エネルギー値を算出する音高別エネルギー統合手段と、
前記統合エネルギー値が高い上位の音高を所定の個数だけ抽出し、抽出された各音高に対応する符号コードに、所定の強さ、所定の発音開始時刻、所定の発音終了時刻のパラメータを設定し、複数の符号コードで構成される音素符号に変換する符号変換手段と、

を有することを特徴とする音素符号変換装置。
【請求項２】
請求項１において、
前記音高別エネルギー統合手段は、日本語カナ文字の「ア」に対応する複数個の「カ」「サ」「タ」「ナ」「ハ」「マ」などの複数の子音音節に対応する音節符号間で対応するエネルギー総和値を音高別に乗算して統合エネルギー値を算出し、
前記符号変換手段は、「Ａ」などの共通する母音音素の音素符号に変換することを特徴とする音素符号変換装置。
【請求項３】
請求項１において、
前記音高別エネルギー統合手段は、日本語カナ文字の「ア」「イ」「ウ」「エ」「オ」からなる５種の母音音節に対応する「カ」「キ」「ク」「ケ」「コ」などの複数の子音音節に対応する音節符号間で対応するエネルギー総和値を音高別に乗算して統合エネルギー値を算出し、
前記符号変換手段は、「Ｋ」などの共通する子音音素の音素符号に変換することを特徴とする音素符号変換装置。
【請求項４】
請求項３において、
前記音高別エネルギー統合手段は、日本語カナ文字の「ア」「イ」「ウ」「エ」「オ」からなる５種の母音音節に対応する「カ」「キ」「ク」「ケ」「コ」などの複数の子音音節に対応する音節符号間で対応するエネルギー総和値を音高別に乗算する際、
あらかじめ、「ア」「イ」「ウ」「エ」「オ」からなる５種の母音音節に対応する音素符号を決定しておき、
各子音音節のエネルギー総和値の中で、各々対応する母音の前記決定された音素符号を構成する全ての音高に対応するエネルギー総和値に所定の１未満の実数値を乗算することにより縮小するようにしていることを特徴とする音素符号変換装置。
【請求項５】
請求項１から請求項４のいずれかに記載の音素符号変換装置により作成され、日本語カナ文字の各音節を構成する音素に対応して、所定の種類以下の音高を同時にもち、音の強さおよび音の長さが均一の複数の符号コードで成される音素符号を、音素符号を識別する音素符号識別情報と対応付けて記録した音素符号データベースと、
与えられた合成指示データに記載されている音節識別情報をもとに音素符号識別情報に変換し、対応する音素符号を前記音素符号データベースから抽出し、前記音節識別情報に従って、発音の開始および終了を特定する時刻を設定し、母音音素に対応する音素の発音の終了を特定する時刻より所定の無音区間を加えた時刻を後続する音節の発音の開始を特定する時刻として設定することにより合成音声データを生成する音素編集処理手段と、
を有することを特徴とする音声合成装置。
【請求項６】
請求項５において、
前記音素編集処理手段により生成された合成音声データを音声として出力する音声出力手段をさらに有することを特徴とする音声合成装置。
【請求項７】
請求項５または請求項６において、
前記音素編集処理手段により生成された合成音声データを五線譜に変換し、印刷する印刷手段をさらに有することを特徴とする音声合成装置。
【請求項８】
請求項５から請求項７のいずれかにおいて、
前記音素編集処理手段は、前記合成指示データに記載されている音節識別情報が母音音節で、日本語カナ文字の長音であるとき、その音節全体の発音時間を、所定の値だけ増加させることを特徴とする音声合成装置。

【請求項９】
請求項５から請求項７のいずれかにおいて、
前記音素編集処理手段は、前記合成指示データに記載されている音節識別情報が、日本語カナ文字の促音であるとき、当該促音の直後に配置される音節に対応する第１の音素と同一の音素を、当該第１の音素の直前に配置して、各音素の発音の開始を特定する時刻、発音の終了を特定する時刻を設定することを特徴とする音声合成装置。
【請求項１０】
請求項５から請求項７のいずれかにおいて、
前記音素編集処理手段は、前記合成指示データに記載されている音節識別情報が、日本語カナ文字の「ヤ」「ユ」「ヨ」の拗音であるとき、前記拗音の直前の音節については、第１の音素のみを設定し、当該第１の音素の直後に、前記拗音に対応する「ヤ」「ユ」「ヨ」いずれかの音節に対応する２つの音素を配置して、各音素の発音の開始を特定する時刻、発音の終了を特定する時刻を設定することを特徴とする音声合成装置。
【請求項１１】
請求項５から請求項１０のいずれかにおいて、
前記音素編集処理手段が、前記音節識別情報より変換された音素符号識別情報に対応する音素符号を前記音素符号データベースから抽出し、前記音節識別情報に従って、発音の開始および終了を特定する時刻を設定する際、前記無音区間に対して、設定された時間伸縮率を乗算し、前記発音の開始および終了を特定する時刻に対して所定の改変を施すようにしていることを特徴とする音声合成装置。
【請求項１２】
請求項５から請求項１１のいずれかにおいて、
前記音素編集処理手段が、前記音節識別情報より変換された音素符号識別情報に対応する音素符号を前記音素符号データベースから抽出し、前記音節識別情報に従って、発音の開始および終了を特定する時刻を設定する際、設定された音高オフセットパラメータに基づいて、前記音素符号データベースに記録されている前記音素符号が母音の場合、当該音素符号を構成する各符号コードの音高に対して、前記音高オフセットパラメータを加算し、前記合成音声データを構成する全ての母音音素に対応する符号コードの音高に対して所定の改変を施すようにしていることを特徴とする音声合成装置。
【請求項１３】
請求項１２において、
前記合成指示データには各音節ごとに音節識別情報とともに前記音高オフセットパラメータが定義されており、前記音素編集処理手段が、与えられた音節識別情報より変換された音素符号識別情報に対応する音素符号を前記音素符号データベースから抽出し、前記音節情報に従って、発音の開始および終了を特定する時刻を設定する際、前記各音節ごとに定義された音高オフセットパラメータに基づいて、前記音素符号データベースに記録されている前記音素符号が母音の場合、当該音素符号を構成する各符号コードの音高に対して、前記音高オフセットパラメータを加算し、前記合成音声データを構成する全ての母音音素に対応する符号コードの音高に対して、改変を施すようにしていることを特徴とする音声合成装置。
【請求項１４】
請求項１から請求項４のいずれかに記載の音素符号変換装置としてコンピュータを機能させるためのプログラム。
【請求項１５】
請求項５から請求項１３のいずれかに記載の音声合成装置としてコンピュータを機能させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、本発明は人間の声を基にして、電子楽器、楽譜等で利用可能な符号データを

得るための技術に関する。
【背景技術】
【０００２】
従来、人間の声を擬似的に合成する手法は、種々の分野で利用されている。出願人は、人間の声をＰＣＭでデジタル化した後、フーリエ変換を用い、実効強度の大きい周波数に対応する符号コードを取得することにより音声合成を行う技術を提案している（特許文献１～５参照）。
【０００３】
また、出願人は、玩具などに搭載されている性能の低いＭＩＤＩ音源でも再生可能とし、既存の楽譜編集ツールに読み込ませて五線譜に変換すると、判読性のある譜面が得られるようにするために、各音節ごとの符号コード群を音素ごとの符号コード群に変換する技術を提案している（特許文献６参照）。
【先行技術文献】
【特許文献】
【０００４】
【特許文献１】特許第３７９５２０１号公報
【特許文献２】特許第４０３７５４２号公報
【特許文献３】特許第４１５６２６８号公報
【特許文献４】特許第４１３２３６２号公報
【特許文献５】特許第４０６１０７０号公報
【特許文献６】特願２００９－１４３８２５号
【発明の概要】
【発明が解決しようとする課題】
【０００５】
上記特許文献６に記載の発明では、ＭＩＤＩ等の符号コード群として構成される音節符号に対して、各音高ごとに発音開始時刻と発音終了時刻との時間差と符号コードの強さとの積で与えられるエネルギー値の総和であるエネルギー総和値を算出し、複数の音節符号間で、各音高ごとに全てのエネルギー総和値を乗算して統合エネルギー値を算出し、統合エネルギー値が高い上位の音高を抽出して、音素符号を得るようにしたので、実際に人間が発音することにより得られた音節符号を利用して、より少ない数で全ての音節を表現可能な音素符号を得ることが可能となった。
【０００６】
しかしながら上記特許文献６に記載の発明では、同一話者が発した日本語音節７１音を録音した波形音声データ一式を高精細なＭＩＤＩデータに変換し、変換された複数の音節ＭＩＤＩデータどうしを掛け合わせることにより、日本語音素２０音に対応する音素ＭＩＤＩデータを抽出してデータベース化する方法をとっていた。この場合、変換された高精細なＭＩＤＩデータの品質、特に音節における子音音素と母音音素の時間軸上における分離性能が生成される音素ＭＩＤＩデータの品質を左右する。しかし、既提案のＭＩＤＩ符号化方式（特許文献１～５）を今回対象とする短い音声の音節信号に適用すると、子音音素成分と母音音素成分が混在してしまい、先提案の音節ＭＩＤＩデータどうしの掛け合わせでは、明瞭性の良い音素ＭＩＤＩデータを生成することが難しいことが判明した。原因としては、周波数解析における時間分解能が不十分で音声の急峻な周波数変動に追従できておらず、音節成分より子音音素を分離・符号化することが実現できていなかった。時間分解能を向上させる方法として、周波数解析時のフレーム長を小さく設定する方法もあるが、返って周波数解析精度が低下し妥当な符号化データを得ることが難しかった。また、限られた和音の中に低音側の声帯基本振動音など子音音素と母音音素に共通する成分が重複して含まれ、明瞭性の低下につながっていた。
【０００７】
そこで、本発明では、従来と同等な周波数解析精度を維持しながら、解析における時間分解能を向上させ、周波数変動を高精度に抽出した音節符号が得られるとともに、符号デ

ータを基本として音声合成機能を実現する場合において音声再生品質の明瞭性を改善することが可能な音素符号変換装置を提供することを課題とする。

【課題を解決するための手段】

【０００８】

上記課題を解決するため、本発明では、発声された１つの音節を所定のサンプリング周波数でサンプル数Ｊの音節波形信号として取得する音節波形取得手段と、前記音節波形信号を当該サンプリング周波数を維持しながら時間軸上でＱ（Ｑは２以上の整数）倍拡大し、Ｊ×Ｑサンプルの拡大波形信号に変換する音節波形拡大手段と、前記拡大波形信号に対して所定の周波数解析を行い、発音開始時刻、発音終了時刻、音高、強さのデータを備える複数の符号コードで表現した符号コード群を作成する符号コード群作成手段と、前記符号コード群を構成する全ての符号コードの音高をＱ倍に補正し、全ての符号コードの発音開始時刻と発音終了時刻を１／Ｑ倍に補正し、補正された符号コード群で構成される音節符号を作成する音節符号作成手段と、前記音節符号を構成する各符号コードについて、各音高ごとに発音開始時刻と発音終了時刻との時間差と符号コードの強さとの積で与えられるエネルギー値の総和であるエネルギー総和値を算出する音高別エネルギー算出手段と、複数の音節符号間で、各音高ごとに全てのエネルギー総和値を乗算して統合エネルギー値を算出する音高別エネルギー統合手段と、前記統合エネルギー値が高い上位の音高を所定の個数だけ抽出し、抽出された各音高に対応する符号コードに、所定の強さ、所定の発音開始時刻、所定の発音終了時刻のパラメータを設定し、複数の符号コードで構成される音素符号に変換する符号変換手段を有する音素符号変換装置を提供する。

【０００９】

本発明によれば、デジタル化された音響信号の各強度配列を時間軸方向に所定の倍率だけ拡大した後、所定数Ｔ個の強度配列で構成される単位区間ごとに、所定数Ｐ種類の周波数に対応したスペクトル強度を算出し、周波数、時刻を含むＰ個の符号コードを得て、Ｐ個の符号コードの周波数をＱ倍、時刻を１／Ｑ倍に補正するようにしたので、従来と同等な周波数解析精度を維持しながら、解析における時間分解能を向上させ、主として音声信号における周波数変動を高精度に抽出した音節符号が得られ、さらに、得られた音節符号に対して、各音高ごとに発音開始時刻と発音終了時刻との時間差と符号コードの強さとの積で与えられるエネルギー値の総和であるエネルギー総和値を算出し、複数の音節符号間で、各音高ごとに全てのエネルギー総和値を乗算して統合エネルギー値を算出し、統合エネルギー値が高い上位の音高を抽出するようにしたので、実際に人間が発音することにより得られた音節符号を利用して、より少ない数で全ての音節を表現可能な音素符号を得ることが可能となる。

【００１２】

また、本発明では、前記音素符号変換装置により作成され、日本語カナ文字の各音節を構成する音素に対応して、所定の種類以下の音高を同時にもち、音の強さおよび音の長さが均一の複数の符号コードで構成される音素符号を、音素符号を識別する音素符号識別情報と対応付けて記録した音素符号データベースと、与えられた合成指示データに記載されている音節情報をもとに音素符号識別情報に変換し、対応する音素符号を前記音素符号データベースから抽出し、前記音節情報に従って、発音の開始および終了を特定する時刻を設定し、母音音素に対応する音素の発音の終了を特定する時刻より所定の無音区間を加えた時刻を後続する音節の発音の開始を特定する時刻として設定することにより合成音声データを生成する音素編集処理手段を有する音声合成装置を提供する。

【００１３】

本発明の音声合成装置によれば、日本語カナ文字の各音節を構成する音素を音素符号として記録した音素符号データベースを有し、与えられた合成指示データに記載されている音節情報をもとに、対応する音素符号を音素符号データベースから抽出し、音節情報に従って、発音の開始および終了を特定する時刻を設定し、母音音素に対応する音素の発音の

終了を特定する時刻より所定の無音区間を加えた時刻を後続する音節の発音の開始を特定する時刻として設定するようにしたので、音声合成を行うことにより作成される符号コード群は玩具などに搭載されている性能の低いＭＩＤＩ音源で再生可能であるとともに、既存の楽譜編集ツールにより演奏者が楽器演奏で再生可能な判読性のある五線譜に変換することが可能となる。

【発明の効果】

【００１４】

本発明によれば、従来と同等な周波数解析精度を維持しながら、解析における時間分解能を向上させ、周波数変動を高精度に抽出した音節符号が得られるとともに、符号データを基本として音声合成機能を実現する場合において音声再生品質の明瞭性を改善することが可能となるという効果を奏する。

【図面の簡単な説明】

【００１５】

【図１】本発明における音節と音素の基本概念を示す図である。
【図２】本発明における音節と音素の基本概念を示す図である。
【図３】本発明に係る音素符号変換装置の一実施形態を示す構成図である。
【図４】音節録音信号から音節符号を作成する処理の概要を示すフローチャートである。
【図５】時間軸方向の拡大、周波数の増加・時間情報の縮小の概念を示す図である。
【図６】音節と音素の対応関係を記録した変換テーブルを示す図である。
【図７】音節符号から音素符号へ変換する際における符号コード群の変化の様子を示す図である。
【図８】音節符号と、音素符号の構成例を示す図である。
【図９】音素符号記憶部１３に格納された男声の音素符号の例を示す図である。
【図１０】音素符号記憶部１３に格納された男声の音素符号の例を示す図である。
【図１１】音素符号記憶部１３に格納された女声の音素符号の例を示す図である。
【図１２】音素符号記憶部１３に格納された女声の音素符号の例を示す図である。
【図１３】図９、図１０に示した男声の音素符号を五線譜化した例を示す図である。
【図１４】図１１、図１２に示した女声の音素符号を五線譜化した例を示す図である。
【図１５】本発明に係る音声合成装置の一実施形態を示す構成図である。

【発明を実施するための形態】

【００１６】

（１．本発明における音節と音素の基本概念）

以下、本発明の好適な実施形態について図面を参照して詳細に説明する。最初に、本発明における音節と音素の基本概念について説明する。日本語の母音は、２つの特徴的な音声フォルマント成分を含む４つ以上の重音による和音で近似できることが知られている。子音は母音成分に加えて、摩擦音など雑音を表現する和音と母音への経過音を表現する和音の３種の連結された和音が理論上必要であるが、経過音は人間の聴覚上の補間現象（音脈分凝）に委ねれば、最初の雑音と母音の２つの和音に近似できる。そこで、出願人は、日本語の母音は単一の４和音、子音は２連の４和音を基本にして全音素を表現し、これらを時間軸上につなぎ合わせれば音声合成を実現できると考えた。

【００１７】

しかしながら、様々な要因により、ＭＩＤＩ音源で種々の楽器音色を設定して再生しても音声の明瞭性に欠けるという問題があった。本発明では、母音、子音等の音節の構成を従来とは根本的に異なるものとした。具体的には、７１音節それぞれ固有の音節符号とするのではなく、日本語５母音と１５子音に対応する２０の音素符号より１つまたは２つの音素符号を選択し組み合わせることにより、７１音節を表現することとした。

【００１８】

まず、母音音節と子音音節の違いについて説明する。本実施形態では、母音音節は第１音素だけの単独音素とする。子音音節は第１音素と第２音素の２音素構成で第１音素は第２音素に比べ短めにする。なお、本実施形態では、母音音節の第１音素、子音音節の第１

音素、第２音素は、いずれも８和音となっている。

【００１９】

子音音節の第１音素は、「Ｋ、Ｓ、Ｔ、Ｎ、Ｈ、Ｍ、Ｒ、Ｇ、Ｚ、Ｄ、Ｂ、Ｐ、Ｙ、Ｗ」の１４音素のいずれか１つが選択され、子音音節第２音素の、母音音節の第１音素は、「Ａ、Ｉ、Ｕ、Ｅ、Ｏ、ｎ」の６音素のいずれか１つが選択される。子音音節には、濁音（「バ」など）、半濁音（「パ」など）を含み、撥音（はつおん「ン」）は第１音素は音素「Ｎ」、第２音素は音素「ｎ」とする。

【００２０】

本発明では、上述のような構成の子音音節と母音音節を用いて、音声の特徴に応じてさらに多種の態様で合成することを特徴とする。音声の特徴に応じた合成態様の分類については、様々なものが考えられるが、本実施形態では、母音、子音（濁音、半濁音、撥音含む）、長音の母音、長音の子音、促音、拗音の６通りに分類している。

【００２１】

図１、図２は、本実施形態における音節と音素の基本概念を示す図である。図１（ａ）～（ｄ）、図２（ａ）（ｂ）は、上記各分類に対応したものとなっている。図１、図２において、各矩形は、音節または音素を示しており、矩形と矩形の間の空白は無音部分を示している。また、左右方向は時間軸であり、矩形の横幅は、音節の発音時間に対応している。音節の発音時間は、基本的な母音、子音の発音時間を単位区間として設定し、他の分類の音節についても、この単位区間を基準にして定められている。単位区間の具体的な長さは適宜設定することが可能であるが、本実施形態では０．２５秒としている。また、詳しくは後述するが、本実施形態では、音節は、２つのパターンで符号化され、１つの音素で構成される音節と、２つの連続する音素（以下、２連音素という）で構成される音節がある。

【００２２】

図１（ａ）は、母音の場合の合成パターンを示す例である。母音の場合、音節は１つの音素（第１音素）で構成される。図１（ａ）に示すように、第１音素を単位区間分発音した後、単位区間分無音とし、その後、他の音節が続く。図１（ｂ）は、子音の場合の合成パターンを示す例である。子音の場合、音節は第１音素と第２音素の２つの音素で構成される。図１（ｂ）に示すように、単位区間の１／４だけ第１音素を発音し、続いて、単位区間の３／４だけ第２音素を発音する。すなわち、第１音素の発音終了と同時に第２音素の発音を開始し、第１音素の発音開始から第２音素の発音終了までがちょうど単位区間となる。その後、単位区間分無音とし、その後、他の音節が続く。

【００２３】

図１（ｃ）は、長音の母音の場合の合成パターンを示す例である。長音の母音の場合、１つの音素で構成されるが、図１（ｃ）に示すように、通常の母音に比べ、単位区間分発音時間を延ばす。すなわち、第１音素を２単位区間分発音する。その後、単位区間分無音とし、その後、他の音節が続く。図１（ｄ）は、長音の子音の場合の合成パターンを示す例である。長音の子音の場合、２つの音素で構成されるが、図１（ｄ）に示すように、通常の子音に比べ、第２音素の発音時間を単位区間分だけ延ばす。第１音素の発音時間は、通常の子音の場合と同じく、単位区間の１／４である。すなわち、第１音素の発音終了と同時に第２音素の発音を開始し、第１音素の発音開始から第２音素の発音終了までが２単位区間分となる。その後、単位区間分無音とし、その後、他の音節が続く。

【００２４】

図２（ａ）は、促音の場合の合成パターンを示す例である。促音の場合、図２（ａ）に示すように、後続音節である子音の第１音素を、後続音節の直前に発音する。例えば、“ッタ”における“ッ”を合成する場合、後続音節“タ”の第１音素を、“タ”の第１音素の直前に配置する。このとき、促音の発音開始は、先行音素の発音終了から単位区間の３／４だけ経過した時点とする。すなわち、先行音素の発音終了から単位区間の３／４だけ経過した時点から、後続音節の第１音素が単位区間の１／４だけ２回繰り返して発音され、その後、後続音節の第２音素が単位区間の３／４だけ発音されることになる。

【００２５】
図２（ｂ）は、拗音の場合の合成パターンを示す例である。拗音の場合、図２（ｂ）に示すように、直前の子音音節を第１音素のみとする。そして、直前の子音音節の第１音素の発音終了時刻を、拗音に対応する子音の発音開始時刻として、その拗音に対応する子音の第１音素と第２音素を連続して発音する。例えば、“キャ”における“ャ”を合成する場合、先行音節“キ”の第１音素の直後に、拗音“ャ”に対応する子音“ヤ”の第１音素を配置する。このとき、直前の子音音節の第１音素の発音時間は単位区間の１／４とし、拗音に対応する子音の発音時間は、通常の子音の場合と同様、第１音素が単位区間の１／４、第２音素が単位区間の３／４となる。
【００２６】
図１（ａ）～図１（ｄ）、図２（ｂ）においては、先行音節を省略し、図２（ａ）、（ｂ）においては、後続音節を省略したが、基本的に全ての音素の間には単位区間分の無音区間が設定される。この単位区間の具体的長さは前述の通り０．２５秒を標準として設定しているが、適宜変更可能である。また、図１、２に示したものは、基本様式であるため、各音節における各音素の発音時間の比率、発音時間や無音時間の長さは適宜変更することが可能である。例えば、隣接音節との間隔を変動させれば話速を変更できる。
【００２７】
（２．前準備）
次に、従来技術を利用した前準備について説明する。前準備として、人間の声を音節単位でデジタル化する。これは、各音節を人間が実際に発声し、これを録音したものをＰＣＭ等の手法により行う。このとき、話者はネイティブの日本人男性または女性とし、同一人が７１の音節に対してできる限り、ピッチ（音高）と発声区間を揃えて、発声することが望まれる。一般人の話者ではこのように発声を揃えるのは困難であるため、専門のボイストレーニングを受けたアナウンサー・声楽家などに発声してもらうことが望ましい。更に、男性または女性を含む複数の話者により複数のセットの７１音節の録音信号を収集することが望ましい。前準備により、日本語７１音節の録音信号が得られる。この日本語７１音節とは、いわゆる五十音に加え、撥音、濁音、半濁音を含む計７１音である。
【００２８】
（３．音素符号への変換）
次に、音節ごとの録音信号から音素符号への変換について説明する。図３は、本発明に係る音素符号変換装置の一実施形態を示す構成図である。記憶手段１０は、音節録音データ記憶部１１、音節符号記憶部１２、音素符号記憶部１３を有しており、コンピュータに接続されたハードディスク等の外部記憶装置により実現される。音節符号記憶部１１には、上述の前処理により作成された録音信号が音節を特定する音節識別情報と対応付けて記憶されている。音節録音データ記憶部１１には、音節と音素の対応関係を示した変換テーブルが記憶されており、符号化された音節符号が音節識別情報と対応付けて記憶される。音節符号記憶部１２には、音節と音素の対応関係を示した変換テーブルが記憶されており、符号化された音節符号が音節識別情報と対応付けて記憶される。音素符号記憶部１３には、符号化された音素符号が音素符号識別情報と対応付けて記憶される。
【００２９】
処理制御手段２０は、音素符号変換装置全体の処理を統括するものであり、音節波形拡大手段２１、符号コード群作成手段２２、音節符号作成手段２３、音高別エネルギー算出手段２４、音高別エネルギー統合手段２５、符号変換手段２６を有している。処理制御手段２０は、ＣＰＵ、メモリを含むコンピュータ本体であり、音節波形拡大手段２１、符号コード群作成手段２２、音節符号作成手段２３、音高別エネルギー算出手段２４、音高別エネルギー統合手段２５、符号変換手段２６は、専用のプログラムをＣＰＵが実行することにより実現される。符号表示手段３０は、処理制御手段２０により処理される音節符号、音素符号を表示するものであり、液晶ディスプレイ等の表示装置により実現される。
【００３０】
続いて、本実施形態に係る音素符号変換装置の処理動作について説明する。図４は、本

実施形態に係る音素符号変換装置において、音節符号を生成するまでの処理概要を示すフローチャートである。

【００３１】

まず、処理制御手段２０は、処理対象である音節録音信号を、図示しないデータ入力機器から読み込む。音節録音信号は、アナログの録音信号を所定のサンプリング周波数、量子化ビット数でサンプリングしたものであり、本実施形態では、サンプリング周波数４４．１ｋＨｚ、量子化ビット数１６ビットでサンプリングした場合を例にとって以下説明していく。サンプリング周波数４４．１ｋＨｚでサンプリングした場合、音節録音信号は、１秒間に４４１００個のサンプルを有するサンプル列（サンプルの配列）として構成されることになる。

【００３２】

音節録音信号を読み込んだら、音節波形拡大手段２１が、音節録音信号を構成するサンプルを時間軸方向に所定の倍率Ｑ（Ｑは整数）だけ拡大する（Ｓ１）。具体的には、音節録音信号を構成するサンプルの数をＱ倍にする。そして、Ｑ個ごとに、元のサンプルと同じ値のものを配置し、その間の（Ｑ－１）個のサンプルの値としては、両側に位置する元のサンプルの値を用いて線形補間したものを与える。原音節録音信号の各サンプルｊ（ｊ＝０・・・Ｊ－１）についてのサンプル値をｘ（ｊ）とすると、コンピュータは、以下の〔数式１〕に従った処理を実行することにより、拡大後の音節録音信号の各サンプルｊ・Ｑ＋ｋ（０≦ｋ≦Ｑ－１）についてのサンプル値ｘ´（ｊ・Ｑ＋ｋ）を算出する。以下の〔数式１〕において、ｗはｋ／（Ｑ－１）で与えられる０≦ｗ≦１の値をとる実数値とする。

【００３３】

〔数式１〕

ｘ´（ｊ・Ｑ＋ｋ）＝（１－ｗ）・ｘ（ｊ）＋ｗ・ｘ（ｊ＋１）

【００３４】

Ｓ１における処理の結果、デジタル音響信号を構成するＪ個のサンプルは、Ｊ×Ｑ個に拡大される。図５（ａ）にＳ１における拡大処理による波形の変化を示す。図５（ａ）における波形は、サンプルの値をプロットしたものを線分で結んだものであるが、サンプル数が多いため、曲線状に表現されるものである。上記〔数式１〕に従った処理を実行することにより、左側に示したような波形が右側に示したような波形に変化することになる。なお、図５の例では、説明の便宜上Ｑ＝２の場合を示している。

【００３５】

次に、符号コード群作成手段２２が、時間軸方向に拡大されたサンプル上に単位区間を設定する（Ｓ２）。単位区間の長さ（サンプル数Ｔ）は、サンプリング周波数との関係で設定されるが、サンプリング周波数が４４．１ｋＨｚの場合、低域部まで忠実に解析するためには、４０９６サンプル以上必要である。そこで、本実施形態では、１単位区間のサンプル数Ｔ＝４０９６として単位区間を設定している。

【００３６】

単位区間の設定は、特許文献１～５に開示されているように、デジタル音響信号の先頭から順次サンプルを抽出することにより行われる。単位区間は、全てのサンプルを漏らさず設定し、好ましくは、連続する単位区間においてサンプルが重複するように設定する。この場合、各単位区間の先頭の間隔（シフト幅という）は、様々な規則で設定することができる。最も単純なのは、シフト幅を固定、すなわち重複させるサンプル数を一定として設定する手法である。例えば、Ｔ＝４０９６の場合、先頭の単位区間をｊ＝０～４０９５、２番目の単位区間をｊ＝２０４８～６１４３、２番目の単位区間をｊ＝４０９６～８１９１というように、２０４８（＝Ｔ／２）個のサンプルを重複させながら、設定することになる。しかし、時間分解能を向上させるためには、シフト幅を小さくしたいという要望があり、一方、シフト幅を小さくするほど、計算時間が増大するという問題がある。また、シフト幅を必要以上に小さくすると、後述する図４・Ｓ４の単音成分の連結処理において連結条件が満足されなくなり連結処理が適切に機能しなくなる。そこで、音響信号の状

態に合わせて最適なシフト幅を設定するため、本実施形態では、特許文献５に開示したような、ゼロ交差点間隔の粗密または自己相関解析により周波数変化が顕著なゼロ交差点を選別し、このゼロ交差点に位置するサンプルを先頭とする。

【００３７】

ゼロ交差点とは、正負両極性の音響信号と、信号の０レベルとの交差点のことであり、ここでは、音響信号の信号の強度値（振幅）が０となる時刻を示す。ただし、デジタル化した音響信号は、アナログ信号におけるゼロ交差点をサンプルするとは限らない。そのため、実際には、強度値が丁度０になる場合に加え、サンプリング点の強度値が正から負、または負から正に変化した場合に、その前後のサンプリング点のどちらかをゼロ交差点とみなす処理を行う。なお、ゼロ交差点検出のためには、解析対象となる音響信号が正負両極性となっている必要がある。そのため、直流成分を含む音響信号については、直流成分を除去しておく必要がある。直流成分の除去については、周知の種々の手法を適用することができるので、ここでは詳細な説明は省略する。基本的には、ゼロ交差点に位置するサンプルを先頭として単位区間を設定するが、連続する単位区間のシフト幅が一定の範囲に収まるように、ゼロ交差点以外の位置を先頭として単位区間を設定する場合もある。具体的には、最大シフト幅（例えばＴ／２）を上回る場合は、ゼロ交差点以外の位置でも最大シフト幅となる位置を先頭にして単位区間を設定する。逆に、最小シフト幅（例えばＴ／８）を下回る場合は、最小シフト幅を上回るように幾つかのゼロ交差点を飛ばした位置を先頭にして単位区間を設定し、最小シフト幅を上回りかつ最大シフト幅の範囲で該当するゼロ交差点が存在しない場合は、上記と同様に最大シフト幅となる位置を先頭にして単位区間を設定するような補正を行う。

【００３８】

符号コード群作成手段２２は、続いて、設定された各単位区間を対象として周波数解析を実行し、各単位区間のスペクトルを算出する（Ｓ３）。各単位区間のスペクトルの算出は、特許文献１～５に開示されているように、ＭＩＤＩのノートナンバーｎ（０≦ｎ≦１２７）に対応する１２８種の解析周波数ｆ（ｎ）＝４４０・$2^{(n-69)/12}$の要素信号（要素関数）を基本にした一般化調和解析により、１２８個の成分を抽出することにより行う。ノートナンバーｎに対応して解析周波数を設定した場合、周波数が高くなるにつれ、ノートナンバー間の周波数間隔が広くなるため、特に、ｎが６０を超えると解析精度が低下してしまう。そこで、本実施形態では、特許文献３に開示したように、ノートナンバー間をＭ個の微分音に分割した１２８Ｍ個の要素信号ｆ（ｎ，ｍ）＝４４０・$2^{(n-69+m/M)/12}$を用いて解析を行い、１２８Ｍ個の成分を抽出する。後述する図４・Ｓ４においてピッチベンド符号の付加など特殊な符号化を行わない限り、各ノートナンバーにおけるＭ個の微分音の情報は不要であるため、Ｍ個の微分音の成分の合算値を当該ノートナンバーにおける成分として代表させ、結果的に１２８個の成分を抽出する。

【００３９】

符号コード群作成手段２２による具体的な処理手順としては、まず、ノートナンバー分の強度配列Ｅ（ｎ）（０≦ｎ≦１２７）と副周波数配列Ｓ（ｎ）を設定し、初期値を全て０とする。続いて、０≦ｎ≦１２７および０≦ｍ≦Ｍ－１に対して以下の〔数式２〕に従った処理を実行し、Ｅ（ｎ，ｍ）を最大にする（ｎｍａｘ，ｍｍａｘ）を求める。

【００４０】

〔数式２〕

$$A(n,m)=(1/T(n))\cdot\Sigma_{i=0,T(n)-1}x(i)\sin(2\pi f(n,m)i/fs)$$

$$B(n,m)=(1/T(n))\cdot\Sigma_{i=0,T(n)-1}x(i)\cos(2\pi f(n,m)i/fs)$$

$$\{E(n,m)\}^2=\{A(n,m)\}^2+\{B(n,m)\}^2$$

【００４１】

上記〔数式２〕においてＴ（ｎ）は解析フレーム長であり、単位区間Ｔを超えない範囲で要素信号の周期の最大の整数倍になるように設定し、ｋを適当な整数値として、Ｔ（ｎ）＝ｋ／ｆ（ｎ，ｍ）で与える。Ｅ（ｎ，ｍ）を最大にする（ｎｍａｘ，ｍｍａｘ）を用いたｆ（ｎｍａｘ，ｍｍａｘ）が調和信号として選出されることになる。（ｎｍａｘ，ｍ

ｍａｘ）が求められたら、コンピュータは、Ａ（ｎｍａｘ，ｍｍａｘ）およびＢ（ｎｍａｘ，ｍｍａｘ）を用いて、以下の〔数式３〕に従った処理を実行し、サンプル配列ｘ（ｉ）の全ての要素（０≦ｉ≦Ｔ－１）を更新する。

【００４２】

〔数式３〕

ｘ（ｉ）←ｘ（ｉ）－Ａ（ｎｍａｘ，ｍｍａｘ）・sin(２πｆ（ｎｍａｘ，ｍｍａｘ）ｉ／ｆｓ)－Ｂ（ｎｍａｘ，ｍｍａｘ）・cos（２πｆ（ｎｍａｘ，ｍｍａｘ）ｉ／ｆｓ）

【００４３】

上記〔数式３〕においては、ｘ（ｉ）から含有信号を減じる処理を行っている。さらに、以下の〔数式４〕に従った処理を実行し、強度配列Ｅ（ｎ）、副周波数配列Ｓ（ｎ）を更新する。

【００４４】

〔数式４〕

Ｅ（ｎｍａｘ）←Ｅ（ｎｍａｘ）＋Ｅ（ｎｍａｘ，ｍｍａｘ）

Ｓ（ｎｍａｘ）←ｍｍａｘ

【００４５】

コンピュータは、上記〔数式２〕～〔数式４〕の処理を全てのｎ（０≦ｎ≦１２７）に対して実行し、全てのＥ（ｎ）およびＳ（ｎ）の値を決定する。

【００４６】

本実施形態では、処理負荷を軽減するため、Ｍの値については、ノートナンバーに基づいて可変に設定し、例えば解析する周波数間隔が１００Ｈｚ程度になるようにしている。そして、ノートナンバー６０以下は分割せずＭ＝１にする。また、精度は若干落ちるが、初回の〔数式２〕の処理でＳ（ｎ）を決定し、２回目以降の〔数式２〕の処理は、ｍ＝Ｓ（ｎ）に固定して行い、微分音解析を省略するようにしても良い。また、〔数式２〕の処理で、既に同一ノートナンバーに対して副周波数が異なる信号成分が複数回に渡って解析される可能性があるが、Ｅ（ｎ）とＳ（ｎ）に既に値がセットされている場合は、Ｅ（ｎ，ｍ）の最大値の選定候補から除外するようにしても良い。

【００４７】

各単位区間について、スペクトル（１２８個の周波数成分）が算出されたら、音節符号作成手段２３が、周波数情報と、各周波数に対応するスペクトル強度、および単位区間の開始と終了を特定可能な時間情報で構成される符号コードを作成する（Ｓ４）。符号コードの作成にあたり、まず、算出したスペクトルに、各ノートナンバーｎの時刻、時間長の情報を追加し、[開始時刻，時間長，主周波数ｎ，副周波数Ｓ（ｎ），強度Ｅ（ｎ）]で構成される単音成分を作成する。「開始時刻」としては単位区間の先頭の時刻を、デジタル音響信号全体において特定できる情報であれば良く、本実施形態では、単位区間の先頭サンプル（ｉ＝０）に付されたデジタル音響信号全体におけるサンプル番号（絶対サンプルアドレス：ｊに対応）を記録している。この絶対サンプルアドレスをサンプリング周波数（４４１００）で除算することにより、音響信号先頭からの時刻が得られる。時間長は、本実施形態では単位区間ごとに可変で与えられることを特徴とし、直後に後続する単位区間の開始時刻までの差分（後続する単位区間の開始時刻－当該単位区間の開始時刻）で与えられる。

【００４８】

Ｓ２で設定された単位区間ごとに、１２８個の単音成分が作成されるが、さらに、Ｓ４においては、連続する単位区間において単音成分を連結する処理を行う。具体的には、連続する単位区間における同一ノートナンバーの単音成分が、所定の連結条件を満たす場合、２つの単音成分を連結する。連結条件としては、同一の音として連続性を有する状態を適宜設定することができるが、本実施形態では、副周波数を考慮した周波数（主周波数＋副周波数）の差が所定の閾値Ｎｄｉｆ未満で、双方の強度が所定の閾値Ｌｍｉｎ以上で、かつ双方の強度の差が所定の閾値Ｌｄｉｆ未満である場合に、連続性を有するとして、後続の単音成分を前方の単音成分に連結する。ただし、連結後の主周波数、副周波数、強度

は大きい方の単音成分の各値を採用し、時間長は双方の和で与える。連結条件としての具体的な閾値は、本実施形態では、Ｎｄｉｆ＝８／２５[単位：ノートナンバー換算]、Ｌｍｉｎ＝１[単位：１２８段階ベロシティ換算]、Ｌｄｉｆ＝１０[単位：１２８段階ベロシティ換算]としている。連結処理は、符号コードへの変換前に行うものであるため、各閾値は、ノートナンバー、ベロシティに換算したものである。

【００４９】

同一ノートナンバーの単音成分の連結は、連結条件を満たす限り、後続する単位区間の単音成分に対して繰り返し行い、最終的に得られた[開始時刻，時間長，主周波数ｎ，副周波数Ｓ（ｎ），強度Ｅ（ｎ）]の単音成分を、符号コードに変換する。符号コードの形式としては、周波数情報と、各周波数に対応するスペクトル強度、および単位区間の開始と終了を特定可能な時間情報を有するものであれば、どのような形式のものであっても良いが、本実施形態では、ＭＩＤＩ形式に変換する。ＭＩＤＩでは、発音開始と、発音終了を別のイベントとして発生するため、したがって、本実施形態では、１つの単音成分を２つのＭＩＤＩノートイベントに変換する。具体的には、「開始時刻」で、ノートナンバーｎのノートオンイベントを発行し、ベロシティ値は強度Ｅ（ｎ）の最大値をＥｍａｘとして、１２８・｛Ｅ（ｎ）／Ｅｍａｘ｝$^{1/4}$で与える。時刻については、Standard MIDI Fileでは、直前イベントとの相対時刻（デルタタイム）で与える必要があり、その時刻単位は任意の整数値で定義でき、例えば、１／１５３６[秒]の単位に変換して与える。そして、「開始時刻」＋「時間長」で特定される終了時刻で、ノートナンバーｎのノートオフイベントを発行する。この際、時間長には、０以上１以下の実数を乗じる。これは、使用するＭＩＤＩ音源の音色にも依存するが、ＭＩＤＩ音源の余韻を考慮して早めにノートオフ指示をするためである。時間長をそのまま用いてもＭＩＤＩ音源の処理上問題はないが、発音の際、後続音と部分的に重なる場合がある。

【００５０】

ＭＩＤＩ符号に変換する際、ＭＩＤＩ音源で処理可能な同時発音数についても考慮するため、同時発音数の調整を行う必要がある。ＭＩＤＩ音源で処理可能な同時発音数が３２である場合、時間軸方向に発音期間中（ノートオン状態）のノートイベントの個数を連続的にカウントし、同時に３２個のノートイベントが存在する箇所が見つかった場合は、各々対になるノートオフイベントを近傍区間内で探索し、各ノートイベント対のベロシティ値とデュレーション値（ノートオフ時刻－ノートオン時刻）の積（エネルギー値）で優先度を評価し、指定和音数（この場合“３２”）以下になるように優先度の低い（エネルギー値の小さい）ノートイベント対を局所的に削除する処理を行う。“局所的に”とは、３２を超えるノートイベントが存在する部分に限りという意味である。この際、ベロシティ値またはデュレーション値のいずれかが所定の下限値より低い場合、優先度に関係なく削除する処理も行う。

【００５１】

さらに、ＭＩＤＩ符号に変換する際、ＭＩＤＩ音源で処理可能なビットレートについても考慮するため、ビットレートの調整を行う必要がある。時間軸方向に、１秒間隔にノートオンまたはノートオフイベントの個数をカウントし、各々の符号長を平均５バイト（４０ビット）とし、ＭＩＤＩ音源で処理可能な最大ビットレートを９０００［ｂｐｓ（ビット／秒）］とすると、１秒間あたりイベント数が９０００／４０＝２２５個を超えている区間が見つかった場合は、その区間に存在するノートオンまたはノートオフイベントと各々対になるノートオフまたはノートオンイベントを近傍区間内で探索し、各ノートイベント対のベロシティ値とデュレーション値（ノートオフ時刻－ノートオン時刻）の積（エネルギー値）で優先度を評価し、指定イベント個数（この場合“２２５”）以下になるように優先度の低い（エネルギー値の小さい）ノートイベント対を局所的に削除する処理を行う。この際、ベロシティ値またはデュレーション値のいずれかが所定の下限値より低い場合、優先度に関係なく削除する処理も行う。

【００５２】

符号コードの作成が行われたら、音節符号作成手段２３は、時間軸方向に拡大して処理

されたことによる変動を是正するため、各符号コードを補正する処理を行う（Ｓ５）。具体的には、まず、全てのノートイベント（ノートオンイベントまたはノートオフイベント）のノートナンバー値に１２・ｌｏｇ$_2$Ｑだけ加算する処理を行う。例えば、Ｑ＝４の場合、２４半音（２オクターブ）だけ全体的に音高を上げる。この処理は、Ｓ１においてサンプル数をＱ倍したことにより周波数が１／Ｑになっているため、周波数をＱ倍にして元の状態に戻すために行う。この補正によりノートナンバーが規格値上限の１２７を超えるノートナンバーをもつ符号コードは削除する。具体的には補正前のノートナンバーが１２８－１２・ｌｏｇ$_2$Ｑ以上の符号コードが削除される。

【００５３】

続いて、全てのノートイベントの時刻（ノートオン時刻またはノートオフ時刻）に１／Ｑを乗算する。これにより、ＭＩＤＩ符号全体の演奏時間、および各ノートイベントの発音時間が１／Ｑに縮小される。この処理は、Ｓ１においてサンプル数をＱ倍したことにより全体の演奏時間がＱ倍になっているため、時刻を１／Ｑにして元の状態に戻すために行う。この処理を行うと、時間あたりのノートイベント数がＱ倍に増大するため、上記Ｓ４で実行したビットレートの調整を再度実行する。

【００５４】

Ｓ５における処理の結果、周波数（音高）はＱ倍になるとともに、時間情報は１／Ｑになる。Ｓ５の補正処理によるＭＩＤＩイベント（ＭＩＤＩ符号のノートイベント）の変化の様子を図５（ｂ）に示す。図５（ｂ）においては、Ｑ＝２の場合のＭＩＤＩイベントの変化を、音符により示している。Ｓ５の補正処理により左側の“ミ”の音符は、右側では１オクターブ高い（周波数が２倍）“ミ”の音符に変化している。一方、左側の四分音符が、右側では時間的に１／２の八分音符に変化している。このようにして得られた音節符号は、音節識別情報と対応付けられて音節符号記憶部１２に記憶される。

【００５５】

また、上記の例では、ビットレートの調整をＳ４、Ｓ５の２回行っているが、これらについては、最終的に少なくとも１回ずつ行われていれば良い。また、上記実施形態では、Ｓ２～Ｓ４の処理について好ましい処理例について具体的に説明したが、これらの処理については、本発明の趣旨を逸脱しない範囲で、公知の特許文献１～５に開示した技術を用いることができる。

【００５６】

音節符号が得られたら、７１の音節単位で構成される音節符号を基に、２０の音素符号に変換する。ここで、音節と音素の対応関係を示した変換テーブルを図６に示す。図６において、カタカナで示す「ア」「イ」・・・の７１音は音節であり、アルファベットで示す「Ａ」「Ｉ」は音素である。母音音素「Ａ」・・・「Ｏ」は水平方向の１３音節のＡＮＤ演算で決定され、子音音素「Ｋ」・・・「Ｐ」は垂直方向の５音節のＡＮＤ演算で決定される。図６に示した横長の枠は、母音音素「Ａ」が１３個の音節「ア」・・・「パ」で決定されることを示しており、縦長の枠は、子音音素「Ｋ」が５個の音節「カ」・・・「コ」で決定されることを示している。なお、子音音素「Ｙ」は３個の音節「ヤ」「ユ」「ヨ」で決定され、子音音素「Ｗ」は２個の音節「ワ」「ヲ」で決定され、音素「ｎ」は１個の音節「ン」で決定される。

【００５７】

処理制御手段２０は、図６に示した変換テーブルを参照し、関連する複数の音節符号を読み込む。例えば、音素符号「Ｋ」を得るために、音節符号記憶部１１から「カ」「キ」「ク」「ケ」「コ」の５個の音節符号を読み込む。

【００５８】

処理制御手段２０が、５個の音節符号を読み込んだら、音高別エネルギー算出手段２４は、各音節符号単位で、構成する全ての符号コードを対象にして、音高（ＭＩＤＩの場合、ノートナンバー）別に、エネルギー総和値を算出する。エネルギー総和値は、各音高における音の強度（ＭＩＤＩの場合、ベロシティ）×発音時間（ＭＩＤＩの場合、デュレーション：ノートオフ時刻－ノートオン時刻）により算出する。ここで、エネルギー総和値

の“総和”とは、１つの音節符号内において、同一音高で２つ以上の符号コードが存在する場合、各符号コードについて総和をとることを意味している。

【００５９】

各符号単位で、全音高についてエネルギー総和値が算出されたら、次に、音高別エネルギー統合手段２５が、５個の音節符号間で、同一の音高別にエネルギー総和値を乗算し、統合エネルギー値を得る。乗算であるため、５個の音節符号のうち、いずれか１つの音節符号に存在しない音高成分については、“０”となる。したがって、統合エネルギー値を求めることにより、５個の音節符号に共通の成分のみが残ることになる。例えば、音素符号「K」を構成する音高成分は、音節符号「カ」「キ」「ク」「ケ」「コ」に共通に含まれている音高成分でなければならない。

【００６０】

音高別に統合エネルギー値が算出されたら、符号変換手段２６が、統合エネルギー値が上位の音高を指定和音数（例：８個）を超えないように選出する。

【００６１】

指定和音数は、事前に設定されるものであり、自由に設定することができるが、本実施形態では、上述のように“８”としている。したがって、本実施形態では、３２個の符号コードから８個の符号コードが、符号変換手段２３により抽出されることになる。さらに、符号変換手段２６は、元の各音節符号を構成する複数の符号コードのうち、最も早い区間開始時刻、最も遅い区間終了時刻を、それぞれ選出された全ての音高の区間開始時刻、区間終了時刻として設定し、選出された音高のベロシティを規定値に設定する。ベロシティの規定値については、ベロシティが“０”～“１２７”の値を取り得るため、本実施形態では、その最大の“１２７”としている。

【００６２】

音高別エネルギー算出手段２４、音高別エネルギー統合手段２５、符号変換手段２６による処理前と処理後の音素符号の変化の様子を図７に示す。図７において、横軸は時間、縦軸は周波数（ノートナンバー）に対応している。グラフ内に配置された矩形は符号コードを示しており、横方向の長さは横軸に従って時間的長さを示しているが、縦方向の長さは縦軸とは異なり、周波数ではなく強度（ベロシティ）を示している。

【００６３】

実際には、図６に示したように、１つの音素は、１３個の音節、または５個の音節により得られる（例外として、音素Ｙは３個の音節、音素Ｗは２個の音節、音素ｎは１個の音節）が、図７の例では、説明の便宜上２つの音節符号α、βを用いて、音素γを抽出する場合を示している。図７（ａ）、（ｂ）は、それぞれ音高別エネルギー算出手段２１による処理前の音節符号α、βを構成する符号コード群を示したものである。上述のように、本実施形態では、同一時刻において３２個の符号コードで音節符号を構成し、指定和音数は８に設定するのが一般的であるが、図７（ａ）では、説明の都合上、同一時刻において符号コードは最大６個となっており、指定和音数は４に設定している場合を示している。また、各符号コードを示す矩形の横方向および縦方向の長さからわかるように、各符号コードの再生時間（終了時刻－開始時刻）および強度も異なっている。

【００６４】

音高別エネルギー算出手段２４、音高別エネルギー統合手段２５による処理後は、全ての音高についての統合エネルギー値が得られる。統合エネルギー値は、図７（ａ）（ｂ）に示されるような、音高別エネルギー算出手段２４により音高ごとに算出される発音開始時刻と発音終了時刻との時間差と符号コードの強さ（ベロシティ）との積の同一音高における総和値であるエネルギー総和値に対して、音節符号αとβとの間で対応する音高におけるエネルギー総和値同士を乗算したものであるので、図で表現するのは難しいが、イメージ的には、図７（ｃ）に示すように、音高に対応して、その統合エネルギー値（図７（ｃ）では、図面上下方向の幅で表現）が得られる。そして、符号変換手段２６により、図７（ｄ）に示すように、統合エネルギー値が大きい音高が４つ選出される。さらに、この後、符号変換手段２６により、各音高に対応する符号コードの強度値は、上述のような規

定値に設定される。「カ」「キ」「ク」「ケ」「コ」の５個の音節符号（最大３２の音高）と、「Ａ」「Ｉ」「Ｕ」「Ｅ」「Ｏ」「Ｋ」の６個の音素符号（最大８の音高）の構成例を図８に示す。

【００６５】

図６に示したように、１つの母音の音素符号は、１３個の音節符号の各音高におけるエネルギー総和値同士を乗算することにより決定され、１つの子音の音素符号は一部の例外を除き５個の各音高におけるエネルギー総和値同士を乗算することにより決定され、基本的にはこれら２０種の音素符号を決定する順位は問わない。そうすると、母音の音素符号と子音の音素符号との間で、ある音高が重複して選出される場合が生じる。即ち、８音など限られた音高で構成される子音の音素と母音の音素が音声合成時に時間的に連続して同一音高が再生される場合が生じ、明瞭性の点で好ましくない。そこで、本願では、子音の音素と母音の音素の符号間で、できるだけ同一の音高が重複して含まれないようにする方法を提案する。そのためには、先に、１３個の音節符号の各音高におけるエネルギー総和値同士を乗算することにより５種類の母音の音素符号を決定する。続いて、５個の母音「ア」「イ」「ウ」「エ」「オ」のいずれかを含む５個の子音の音節符号（例えば、「カ」「キ」「ク」「ケ」「コ」）の各音高におけるエネルギー総和値同士を乗算する前に、各音節符号に含まれる母音の既に決定された音素符号の全ての音高に対応するエネルギー総和値を一定の割合で縮小させる。例えば、エネルギー総和値に１／１０００を乗算させて他の音高のエネルギー総和値に対して相対的に小さな値に改変する。乗算する値は、必ずしも１／１０００とする必要はなく、１未満であれば良い。その後、５個の音節符号の各音高における改変されたエネルギー総和値同士を乗算することにより５種類の母音の音素符号を決定する。このような補正処理を施すことにより、子音の音素と母音の音素の符号間で、同一の音高が重複して選択される確率が低くなる。（ただし、エネルギー総和値の高い音高の種類が元来少ない場合、多少重複して選択されることもある。）

【００６６】

音高別エネルギー算出手段２４、音高別エネルギー統合手段２５、符号変換手段２６は、音節符号記憶部１２に記憶されている各音節符号について処理を行い、得られた各音素符号を音素符号記憶部１３に格納する。音素符号記憶部１３に格納された音素符号の例を図９～図１２に示す。このうち、図９、図１０は男声を符号化したものであり、図１１、図１２は女声を符号化したものである。図９～図１２中、“Ｃ，Ｃ＃，Ｄ，Ｄ＃、Ｅ、Ｆ、Ｆ＃、Ｇ、Ｇ＃、Ａ、Ａ＃、Ｂ”は、“ド、ド＃、レ、レ＃、ミ、ファ、ファ＃、ソ、ソ＃、ラ、ラ＃、シ”の音名の英語表記で、列記されている数字はオクターブ番号を示し、音名とオクターブ番号の対記号でＭＩＤＩ規格のノートナンバーを特定でき、本願ではＭＩＤＩ規格ノートナンバーの６９をＡ３と表記する（国際的にはＡ４をＭＩＤＩ規格ノートナンバーの６９を示す表記も多数存在する）。音素符号を構成する符号コードが、ＭＩＤＩ規格で定義されている場合、市販の楽譜編集ツールにより五線譜に変換することができる。図９、図１０の男声の音素符号を五線譜化した例を図１３に、図１１、図１２の女声の音素符号を五線譜化した例を図１４にそれぞれ示す。

【００６７】

（４．音声の合成）

次に、得られた音素符号を利用した音声の合成について説明する。図１５は、本発明に係る音声合成装置の一実施形態を示す構成図である。図１５において、音素符号データベース１３ａは、得られた音素符号を、音素符号識別情報と対応付けて記録したものである。音素符号データベース１３ａに格納されている音素符号は、上述の音素符号変換装置により変換され、音素符号記憶部１３に格納されたものと同じである。したがって、上述の音素符号変換装置は、この音素符号データベース１３ａを作成するためのものであるとも言える。また、音素符号データベース１３ａには、各音素符号識別情報と、音節を特定する音節識別情報との対応関係を示した変換テーブルが記録されている。この変換テーブルは、図６に示したものと同じである。合成音声データ記憶手段１４は、音素編集処理手段５０により合成された合成音声データを記憶するものであり、ハードディスク等の記憶装

置により実現される。

【００６８】

音素編集処理手段５０は、合成指示データの内容に従って、音素符号データベース１３ａから対応する音素符号を抽出し、所定の加工を施して合成音声データを生成し、所定の出力先に出力する処理を行う。生成された合成音声データは、設定に従って合成音声データ記憶手段１４、音声出力手段６０、印刷手段７０のうち、１つ以上に出力される。音声出力手段６０は、音素編集処理手段５０から受け取った合成音声データを実際の音声として発音するものであり、ＭＩＤＩ音源を備えたＭＩＤＩ再生装置により実現される。印刷手段７０は、音素編集処理手段５０から受け取った合成音声データを五線譜に変換し、印刷するものであり、五線譜への変換は、公知の変換ソフトウェアを実行することにより実現され、印刷機能は、公知のプリンタ等により実現される。図１５に示した音声合成装置は、現実には、入力機器、外部記憶装置を備え、ＭＩＤＩ再生装置を接続したコンピュータに専用のプログラムを組み込むことにより実現される。

【００６９】

音声合成装置に入力される合成指示データは、音節識別情報を所定の順序で配置したものであり、この音節識別情報は、音節を識別することができるものであれば、どのような形式であっても良い。本実施形態では、音節識別情報として、音節に対応する文字コードを記録したテキストデータを用いている。この場合、音素符号データベース１３ａ内の変換テーブルには、音節識別情報に対応する文字コードと音素符号識別情報が対応付けて記録されている必要がある。

【００７０】

続いて、図１５に示した音声合成装置の処理動作について説明する。まず、合成指示データを音声合成装置に入力する。音声合成装置は、合成指示データを読み込むと、音素編集処理手段５０が合成指示データ内を先頭の音節識別情報から順に合成処理していく。具体的には、音素編集処理手段５０は、合成指示データ内の音節識別情報で音素符号データベース１３ａ内の変換テーブルを参照して、音素符号識別情報を取得し、その音素符号識別情報に対応する音素符号を抽出する。

【００７１】

そして、抽出した音素符号が母音音素１つだけである場合は、母音音節であるので、先行する音節のノートオフ時刻の０．２５秒後をノートオン時刻として設定し、その０．２５秒後をノートオフ時刻とし、ノートナンバー、ベロシティは音素符号データベース１２ａに記録されていた値そのものとするＭＩＤＩイベントを作成する。ただし、ノートナンバーについてはオプション的に別途ユーザにより指示される音高オフセットパラメータに基づいて適宜上下され、ピッチ変換を行えるようにしてある。

【００７２】

抽出した音素符号が２つであり、それが子音音素と母音音素である場合は、子音音節であるので、先の子音音素符号について、前の音節のノートオフ時刻の０．２５秒後をノートオン時刻として設定し、単位区間の１／４、すなわち０．０６２５秒後をノートオフ時刻とする。そして、後の母音音素符号について、先の子音音素符号のノートオフ時刻をノートオン時刻として設定し、単位区間の３／４、すなわち０．１８７５秒後をノートオフ時刻とする。子音音節の場合も、母音音節の場合と同様、ノートナンバー、ベロシティは音素符号データベース１２ａに記録されていた値そのものとするが、ノートナンバーについてはオプション的に別途ユーザにより指示される音高オフセットパラメータを加算することにより適宜上下され、ピッチ変換を実現することができる。また、上記の０．２５秒、０．０６２５秒、０．１８７５秒という時間数値はあくまで基準値であり、別途ユーザにより指示される時間伸縮パラメータを乗算することにより適宜伸縮され、話速変換を実現することができる。

【００７３】

音節識別情報が長音を示すものであった場合（音節識別情報を文字コードで記録したときは、“ー”に対応する文字コードであった場合）、その直前の音節識別情報とともに２

つの音節識別情報で１つの長音の音節を特定する。例えば、例えば、音節識別情報が“ア”と“ー”が連続した場合、２つの音節識別情報“アー”により、長音の母音音節であると判断する。音節識別情報が“カ”と“ー”が連続した場合、２つの音節識別情報“カー”により、長音の子音音節であると判断する。長音の場合、長音の母音音節と長音の子音音節で若干異なる。長音の母音音節の場合、ノートオン時刻からノートオフ時刻の間隔を０．５秒に増加して設定する。長音の子音音節の場合、先の子音音素符号については、長音でない通常の場合と同様、ノートオン時刻からノートオフ時刻の間隔を０．０６２５秒にして設定し、後の母音音素符号についてノートオン時刻からノートオフ時刻の間隔を０．４３７５秒にして設定する。したがって、長音の場合、音節全体の発音時間は、母音音節、子音音節ともに同じ０．５秒となる。子音音節については、第１音素の発音時間は、長音でない通常の場合と同じ０．０６２５秒であるが、第２音素の発音時間が、長音でない通常の場合と比べて長くなる。尚、上記の０．５秒、０．２５秒、０．４３７５秒、０．５秒、０．０６２５という時間数値も同様にあくまで基準値であり、別途ユーザにより指示される時間伸縮パラメータを乗算することにより適宜伸縮され、話速変換を実現することができる。

【００７４】

促音の場合、その直後の音節の第１音素と同じものを、直後の音節の第１音素の直前に加える。第１音素の発音時間は０．０６２５秒であるため、先行する音節の発音終了時刻から０．１８７５秒後に促音のノートオン時刻を設定することになり、促音のノートオフ時刻と、直後の音節の第１音素のノートオン時刻が同一となる。尚、上記の０．０６２５秒、０．１８７５秒という時間数値も同様にあくまで基準値であり、別途ユーザにより指示される時間伸縮パラメータを乗算することにより適宜伸縮され、話速変換を実現することができる。

【００７５】

拗音の場合、直前の子音の第１音素の直後に加える。したがって、直前の子音の第１音素のノートオフ時刻と、拗音のノートオン時刻が同一となるように設定する。拗音の音節の構成自体は子音と同じであるので、拗音の第１音素のノートオフ時刻および第２音素のノートオン時刻は、第１音素のノートオン時刻の０．０６２５秒後であり、拗音の第２音素のノートオフ時刻は、そのノートオン時刻の０．１８７５秒後となる。尚、上記の０．０６２５秒、０．１８７５秒という時間数値も同様にあくまで基準値であり、別途ユーザにより指示される時間伸縮パラメータを乗算することにより適宜伸縮され、話速変換を実現することができる。

【００７６】

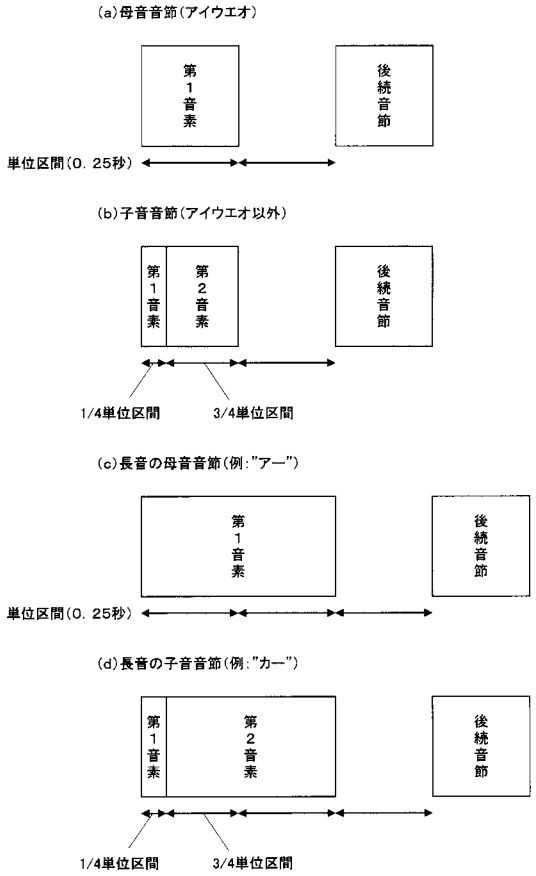
音素編集処理手段５０は、読み込んだ合成指示データ内の音節識別情報単位で音素の合成処理を行っていき、処理が終わった音節単位で順に、合成音声データ（ＭＩＤＩデータ）を、音声出力手段６０に渡していく。音声出力手段６０は、音素編集処理手段５０から受け取ったＭＩＤＩデータを順に再生していく。以上のようにして、音声合成装置は、読み込んだ合成指示データに従って音声の再生が可能となる。

【００７７】

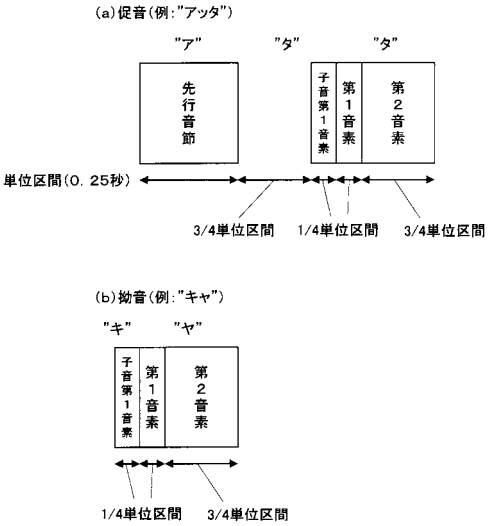
五線譜として出力する場合は、合成音声データを印刷手段７０により五線譜データに変換した後、印刷出力する。また、上記の例のように、合成指示データに従って音声合成をリアルタイムで行い、音声再生したり、五線譜出力することも可能であるが、この音声合成装置では、音素編集処理手段５０による処理結果であるＭＩＤＩデータを合成音声データ記憶手段１３に蓄積し、別途このＭＩＤＩデータをＭＩＤＩ再生装置により音声再生するようにしても良い。ＭＩＤＩデータを記憶装置に蓄積する方法としては、ＳＭＦ(Standard MIDI File)形式ファイルを用いると、市販の種々の音楽関係ソフトウェアに渡すことができ、作成されたＭＩＤＩデータからは、市販の楽譜作成ツールを用いて、楽譜を作成することができる。この場合、楽譜は、ＳＭＦ形式に記録されていた音素符号を基にして作成される。そして、作成された楽譜を印刷装置から出力すれば、読みやすい楽譜として、楽器演奏の際に利用することができる。

【００７８】
上述の通り、音素編集処理手段５０は、合成指示データ内の音節識別情報で音素符号データベース１３ａから対応する音素符号を抽出し、ＭＩＤＩイベントを作成する際、そのノートナンバーについては音素符号データベース１３ａに収録されている当該音素符号を構成する各音符のノートナンバーに対して、オプション的に別途ユーザにより指示される音高オフセットパラメータを加算し適宜上下させ、ピッチ変換を行えるようにしてある。この場合は、合成音声データ全体のピッチを上下させるのではなく、母音音素に限定して上下させるようにする。また、合成指示データ内の音節識別情報とともに音高オフセットパラメータを音節ごとに定義すれば、各音節ごとに構成される母音音素のピッチを個別に上下させることもできる。すなわち、あらかじめ作成した旋律の隣接音符間での音高変化（音程情報）を、合成指示データ内の音節識別情報とともに定義される音高オフセットパラメータとして与えれば、歌声合成を実現することができる。
【産業上の利用可能性】
【００７９】
本発明は、イベントや余興目的に行われる人間の音声再生を模倣した音楽作品制作・作曲の支援産業に利用することができる。また、エンターテインメント分野において、電子楽器を主体とした玩具（ロボット、ぬいぐるみを含む）、玩具型のアコースティック楽器（室内装飾用のミニチュアピアノ）、オルゴール、携帯電話の着信メロディ等の音階再生媒体に対して音声合成機能を付加する産業に利用することができる。また、ＳＭＦ（Ｓｔａｎｄａｒｄ　ＭＩＤＩ　Ｆｉｌｅ）等によるＭＩＤＩ音楽コンテンツ配布時における著作権保護等の産業に利用することができる。
【符号の説明】
【００８０】
１０・・・記憶手段
１１・・・音節録音データ記憶部
１２・・・音節符号記憶部
１３・・・音素符号記憶部
１３ａ・・・音素符号データベース
１４・・・合成音声データ記憶手段
２０・・・処理制御手段
２１・・・音節波形拡大手段
２２・・・符号コード群作成手段
２３・・・音節符号作成手段
２４・・・音高別エネルギー算出手段
２５・・・音高別エネルギー統合手段
２６・・・符号変換手段
３０・・・符号表示手段
５０・・・音素編集処理手段
６０・・・音声出力手段
７０・・・印刷手段

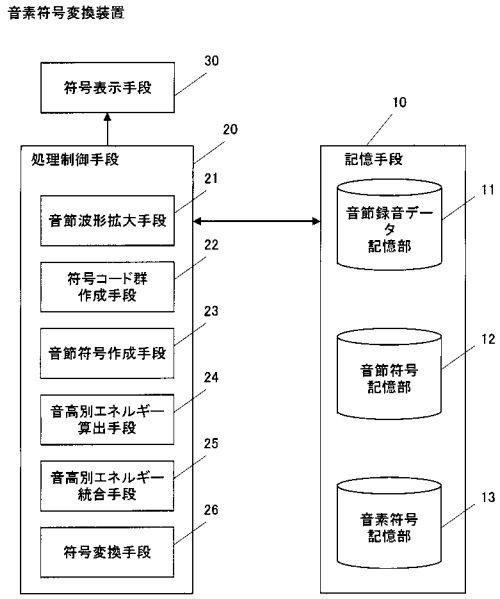
【図１】


(a)母音音節(アイウエオ)
第1音素
後続音節
単位区間(0.25秒)
(b)子音音節(アイウエオ以外)
第1音素
第2音素
後続音節
1/4単位区間
3/4単位区間
(c)長音の母音音節(例:"アー")
第1音素
後続音節
単位区間(0.25秒)
(d)長音の子音音節(例:"カー")
第1音素
第2音素
後続音節
1/4単位区間
3/4単位区間

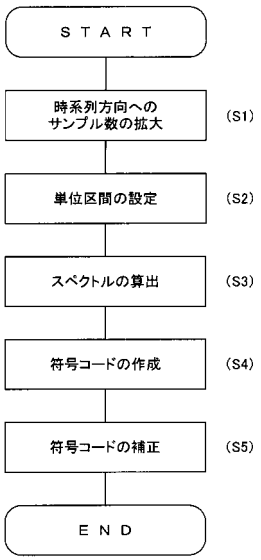
【図２】


(a)促音(例:"アッタ")
"ア"
"タ"
"タ"
先行音節
子音第1音素
第1音素
第2音素
単位区間(0.25秒)
3/4単位区間
1/4単位区間
3/4単位区間
(b)拗音(例:"キャ")
"キ"
"ヤ"
子音第1音素
第1音素
第2音素
1/4単位区間
3/4単位区間

【図３】


音素符号変換装置
符号表示手段
30
処理制御手段
20
21
音節波形拡大手段
22
符号コード群作成手段
23
音節符号作成手段
24
音高別エネルギー算出手段
25
音高別エネルギー統合手段
26
符号変換手段
10
記憶手段
11
音節録音データ記憶部
12
音節符号記憶部
13
音素符号記憶部

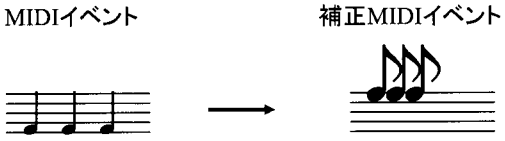
【図４】


START
時系列方向へのサンプル数の拡大
(S1)
単位区間の設定
(S2)
スペクトルの算出
(S3)
符号コードの作成
(S4)
符号コードの補正
(S5)
END

【図５】

(a)時間軸方向にQ倍に拡大

(b)周波数をQ倍、時間軸を1／Q



【図６】

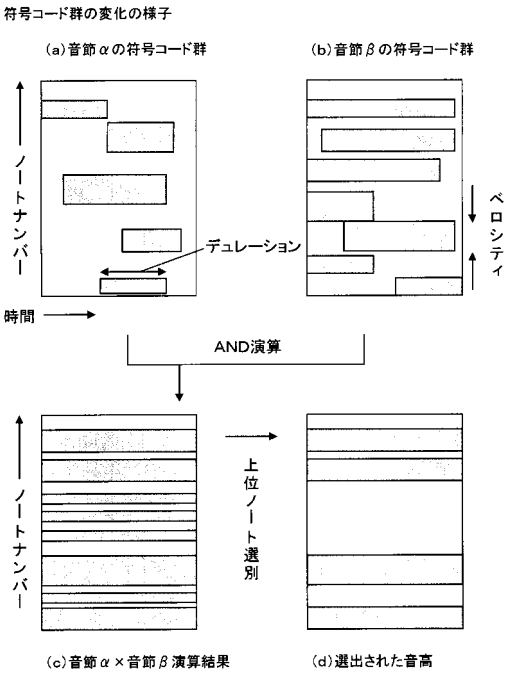
変換テーブル

		K	S	T	N	H	M	R	G	Z	D	B	P	Y	W
A	ア	カ	サ	タ	ナ	ハ	マ	ラ	ガ	ザ	ダ	バ	パ	ヤ	ワ
I	イ	キ	シ	チ	ニ	ヒ	ミ	リ	ギ	ジ	ヂ	ビ	ピ		
U	ウ	ク	ス	ツ	ヌ	フ	ム	ル	グ	ズ	ヅ	ブ	プ	ユ	
E	エ	ケ	セ	テ	ネ	ヘ	メ	レ	ゲ	ゼ	デ	ベ	ペ		
O	オ	コ	ソ	ト	ノ	ホ	モ	ロ	ゴ	ゾ	ド	ボ	ポ	ヨ	ヲ
n					ン										

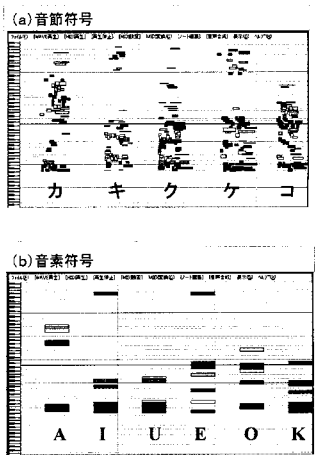
音節：カタカナ（71個）
音素：アルファベット（20個）

【図７】

符号コード群の変化の様子



【図８】



【図９】

日本語音素MIDIコード・データベース例（男声）1

[A]	[I]	[U]	[E]	[O]
F5	A#6	D3	A#3	F#4
E5	C#3	C#3	A3	F4
D#5	A#2	A#2	G#3	E4
A4	D2	A2	E3	G#3
G#4	C#2	D2	G#2	G3
C2	C2	C#2	D2	F#3
B1	B1	C2	C#2	C2
A#1	A#1	B1	A1	B1

[K]	[S]	[T]	[N]	[H]
F4	B6	D#4	C4	D4
A#3	A6	F3	D#3	A3
C3	C4	D#3	G#2	F3
B2	B3	B2	G2	E3
A#2	F3	G#2	F#2	D#3
G2	F2	F#2	F2	F#2
C#2	E2	E2	E2	E2
A1	D#2	A1	A1	A1

【図１０】

日本語音素MIDIコード・データベース例（男声）2

[M]	[R]	[G]	[Z]	[D]
B3	B3	C#4	C#4	D4
D#3	D#3	B3	C4	F3
C3	C3	F3	C3	C3
G2	B2	C3	B2	G#2
F#2	G#2	B2	G2	G2
F2	G2	G2	F#2	F2
E2	F2	F#2	E2	E2
D#2	A1	E2	D#2	A1

[B]	[P]	[Y]	[W]	[n]
C4	C4	C#4	E4	D3
B3	A#3	B3	D4	B2
F3	A3	A3	C#4	A2
D#3	F3	F3	B3	G#2
C3	G2	C3	G#3	D2
B2	E2	G#2	G3	C#2
G2	D#2	D#2	C2	C2
E2	A1	A1	A#1	B1

【図１１】

日本語音素MIDIコード・データベース例（女声）1

[A]	[I]	[U]	[E]	[O]
B5	A6	E5	G6	B4
A#5	G#6	D#5	D#6	A#4
F#5	G6	D#4	B4	A4
D#5	F#6	C#4	F4	G#4
C#5	G3	D#3	E4	D4
C5	F3	D3	D4	C#4
B4	E3	C#3	C#4	A#3
G#4	D#3	A#2	C4	A3

[K]	[S]	[T]	[N]	[H]
D#5	G#6	G#6	C3	F4
B4	E4	D6	B2	E4
D#4	D#4	G#5	A#2	D#4
C4	B3	F4	A2	B3
A3	E3	E4	G#2	A#3
D#3	D3	D#4	F2	E3
D3	C#3	E3	C2	D#3
C#3	B2	G#2	A#1	B2

【図１２】

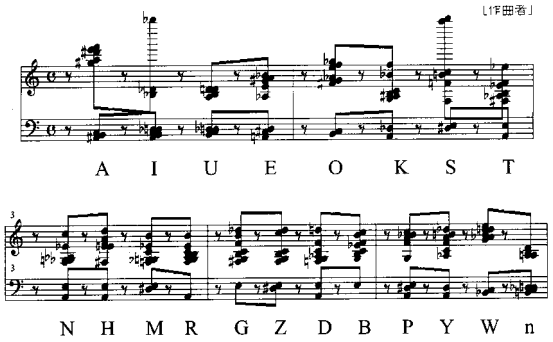
日本語音素MIDIコード・データベース例（女声）2

[M]	[R]	[G]	[Z]	[D]
F#4	G#6	G6	A6	G#6
C3	D#4	F#6	G#6	D#4
B2	C4	E4	G#5	C4
A#2	B3	D4	D3	B3
G#2	C3	C4	C3	D3
F#2	B2	B3	A2	C3
F2	A2	D#3	G#2	G2
C#2	G#2	B2	C2	C2

[B]	[P]	[Y]	[W]	[n]
G4	E5	D#6	D#5	B5
D#4	E4	F#4	D5	A#5
B3	D#4	E4	B4	C4
B2	D4	C4	A#4	A#3
A2	B3	B3	E4	D#3
G#2	A#3	E3	D#4	D3
G2	E3	A2	D4	C3
B1	D#3	G#2	C4	B2

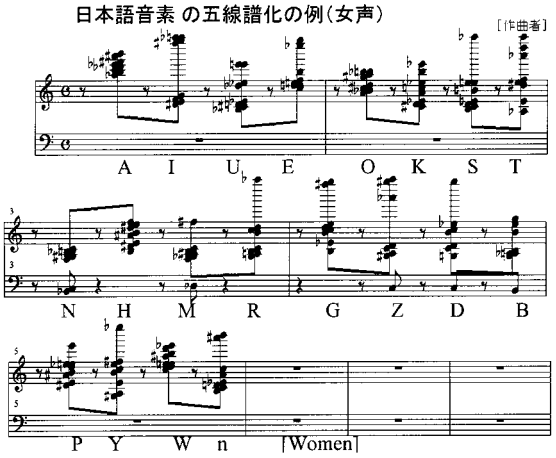
【図１３】

日本語音素 の五線譜化の例（男声）

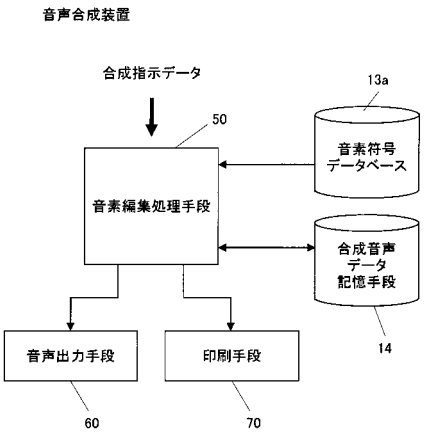


【図１４】

日本語音素 の五線譜化の例（女声）



【図１５】



フロントページの続き

(72)発明者　茂出木　敏雄
東京都新宿区市谷加賀町一丁目１番１号　大日本印刷株式会社内

審査官　山下　剛史

(56)参考文献　特開平４－３４９４９７（ＪＰ，Ａ）
特開２００４－２９４８１６（ＪＰ，Ａ）
特開２００６－１３９１５８（ＪＰ，Ａ）
特開平１０－２４７０９９（ＪＰ，Ａ）
茂出木敏雄，"音声ＭＩＤＩコードを用いたＳＭＦファイルへの情報埋め込み手法"，電子情報通信学会２００９年総合大会講演論文集，２００９年　3月，情報・システム２，pp.S-35～S-36

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１３／００－１３／１０，１９／００，２５／００－２５／９３
Ｇ１０Ｇ　３／０４
Ｇ１０Ｈ　１／００