

【公報種別】特許法第17条の2の規定による補正の掲載  
 【部門区分】第6部門第2区分  
 【発行日】令和5年12月1日(2023.12.1)

【国際公開番号】WO2022/190502  
 【出願番号】特願2023-505112(P2023-505112)

【国際特許分類】

G 1 0 L 1 3 / 0 0 ( 2 0 0 6 . 0 1 )

G 1 0 L 1 3 / 0 2 ( 2 0 1 3 . 0 1 )

G 1 0 K 1 5 / 0 4 ( 2 0 0 6 . 0 1 )

10

【 F I 】

G 1 0 L 1 3 / 0 0 1 0 0 Y

G 1 0 L 1 3 / 0 2 1 3 0 Z

G 1 0 K 1 5 / 0 4 3 0 2 D

【手続補正書】

【提出日】令和5年9月8日(2023.9.8)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】全文

20

【補正方法】変更

【補正の内容】

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音生成装置およびその制御方法、プログラム、電子楽器に関する。

【背景技術】

【0002】

電子鍵盤装置などの電子楽器において、楽器音などを想定した電子音を生成する以外にも、歌唱音を合成して生成することが行われている。このような歌唱音（実際の歌唱と区別して、以下、合成歌唱音という）は、例えば、歌詞などの文字に応じた音声素片を組み合わせつつ、指定された音高になるように波形を合成することによって、その文字を発音したような合成音が生成される。従来は、予め用意した楽譜（シーケンスデータ等）と文字とを組み合わせ、合成歌唱音を生成する技術が用いられていたが、特許文献1、2のように、電子鍵盤装置への演奏操作に応じてリアルタイムに合成歌唱音を生成していく技術も開発されている。

30

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2016-206496号公報

40

【特許文献2】特開2014-98801号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

従来の歌唱音合成装置では、電子鍵盤装置の押鍵に応じて自動的に1文字ずつあるいは1音節ずつ進める場合、ミスタッチや装飾音等があると、歌詞の位置が演奏よりも先に進んでしまうことがあった。歌詞の位置が演奏よりも先に進んでしまった場合、歌詞の位置と演奏とが合っておらず、不自然な合成歌唱音となってしまうという問題があった。

【0005】

そこで、本発明は、リアルタイム演奏で歌唱音を発声させる際に、自然な合成歌唱音を

50

生成することを目的の一つとする。

【課題を解決するための手段】

【0006】

上記目的を達成するために本発明によれば、発声すべき複数の文字が時系列に並べられ、第1文字および当該第1文字の後の第2文字を少なくとも含む第1歌詞データを取得する第1の取得部と、発声の開始指示を取得する第2の取得部と、前記第2の取得部により前記発声の開始指示が取得された場合において、前記発声の開始指示が第1条件を満たす場合には、前記第1歌詞データにおける前記第1文字に対応する第1発声に基づくオーディオ信号を生成する指示を出力し、前記発声の開始指示が前記第1条件を満たさない場合には、前記第1歌詞データにおける前記第2文字に対応する第2発声に基づくオーディオ信号を生成する指示を出力する制御部と、を有し、前記第1歌詞データにおいて、所定の時間軸により、各文字の発声の開始時刻および停止時刻が規定され、前記発声の開始指示の取得時刻が、前記所定の時間軸における、前記第1発声を停止する第1時刻と前記第2発声を開始する第2時刻との間に位置する第3時刻より前である場合に、前記発声の開始指示が前記第1条件を満たす、音生成装置が提供される。

10

【発明の効果】

【0007】

本発明によれば、リアルタイム演奏で歌唱音を発声させる際に、自然な合成歌唱音を生成することができる。

【図面の簡単な説明】

20

【0008】

【図1】本発明の一実施形態におけるカラオケシステムの構成を示すブロック図である。

【図2】本発明の一実施形態における電子楽器の構成を示すブロック図である。

【図3】本発明の一実施形態における第1歌詞データを説明する図である。

【図4】本発明の一実施形態における音生成処理を説明するフローチャートである。

【図5】指示処理を説明するフローチャートである。

【図6】音生成処理における時刻と音高との関係を表す図である。

【図7】音生成処理における時刻と音高との関係を表す図である。

【図8】本発明の一実施形態における音生成機能を示す機能ブロック図である。

【図9】指示処理を説明するフローチャートである。

30

【図10】音生成処理における時刻と音高との関係を表す図である。

【図11】本発明の一実施形態における第1歌詞データを説明する図である。

【図12】音生成処理における時刻と音高との関係を表す図である。

【図13】音生成処理における時刻と音高との関係を表す図である。

【図14】本発明の一実施形態における第2歌詞データを説明する図である。

【図15】本発明の一実施形態における電子管楽器の構成を示すブロック図である。

【発明を実施するための形態】

【0009】

以下、本発明の一実施形態におけるカラオケシステムについて、図面を参照しながら詳細に説明する。以下に示す実施形態は本発明の実施形態の一例であって、本発明はこれら

40

の実施形態に限定されるものではない。

<実施形態>

[全体構成]

本発明の一実施形態におけるカラオケシステムは、合成歌唱音を生成することができる電子楽器を用いてカラオケをする場合に、対象となる楽曲を指定することによって、リアルタイム演奏で歌唱音を発声させる際に、自然な合成歌唱音を生成する機能を有している。

【0010】

図1は、本発明の第1実施形態におけるカラオケシステムの構成を示すブロック図である。カラオケシステム100は、カラオケ装置1、制御端末2、電子楽器3（音生成装置

50

）、カラオケ用サーバ1000、および歌唱音合成用サーバ2000を備える。この例では、カラオケ装置1、カラオケ用サーバ1000および歌唱音合成用サーバ2000は、インターネットなどのネットワークNWを介して接続されている。この例では、カラオケ装置1は、制御端末2および電子楽器3のそれぞれと、短距離無線通信によって接続されるが、ネットワークNWを介した通信によって接続されてもよい。短距離無線通信とは、例えば、Bluetooth（登録商標）、赤外線通信、LAN（Local Area Network）などを用いた通信である。

#### 【0011】

カラオケ用サーバ1000は、カラオケ装置1においてカラオケを提供するために必要な楽曲データを、曲IDに対応付けて記憶している記憶装置を備える。楽曲データは、カラオケの歌唱曲に関連するデータ、例えば、リードボーカルデータ、コーラスデータ、伴奏データ、およびカラオケ用字幕データなどが含まれている。リードボーカルデータは、歌唱曲の主旋律のパートを示すデータである。コーラスデータは、主旋律に対するハモリなどの副旋律のパートを示すデータである。伴奏データは、歌唱曲の伴奏音を示すデータである。リードボーカルデータ、コーラスデータ、および伴奏データは、MIDI形式で表現されたデータであってもよい。カラオケ用字幕データは、カラオケ装置1のディスプレイに歌詞を表示するためのデータである。

10

#### 【0012】

歌唱音合成用サーバ2000は、歌唱曲に合わせた設定を電子楽器3に対して行うための設定データを、曲IDに対応付けて記憶している記憶装置を備える。設定データには、曲IDに対応する歌唱曲の各パートに対応する歌詞データが含まれている。リードボーカルパートに対応する歌詞データを第1歌詞データという。歌唱音合成用サーバ2000に記憶されている第1歌詞データは、カラオケ用サーバ1000に記憶されているカラオケ用字幕データとは、同じであってもよいし、異なってもよい。すなわち、歌唱音合成用サーバ2000に記憶されている第1歌詞データは、発声すべき歌詞（文字）を規定するデータである点においては同じであるが、電子楽器3において利用しやすい形式に調整されている。例えば、カラオケ用サーバ1000に記憶されるカラオケ用字幕データとして、「こ（ko）」「ん（n）」「に（ni）」「ち（chi）」「は（ha）」という文字列である。これに対し、歌唱音合成用サーバ2000に記憶されている第1歌詞データは、電子楽器3において利用しやすいように「こ（ko）」「ん（n）」「に（ni）」「ち（chi）」「わ（wa）」という実際の発音に合わせた文字列であってもよい。また、この形式としては、例えば、1音で2文字分の歌唱をする場合を識別する情報、フレーズの区切りを識別する情報などを含む場合がある。

20

30

#### 【0013】

カラオケ装置1は、オーディオ信号が供給される入力端子、およびオーディオ信号を音として出力するスピーカを含む。入力端子に入力されるオーディオ信号は、電子楽器3から供給されてもよいし、マイクロフォンから供給されてもよい。

#### 【0014】

カラオケ装置1は、カラオケ用サーバ1000から受信した楽曲データのうち伴奏データからオーディオ信号を再生し、歌唱曲の伴奏音としてスピーカから出力する。入力端子に供給されたオーディオ信号に応じた音が伴奏音に合成されて出力されてもよい。

40

#### 【0015】

制御端末2は、カラオケ装置1に対するユーザの指示（例えば、歌唱曲の指定、音量、トランスポーズ等）を送信するリモコンである。制御端末2は、カラオケ装置1を介して電子楽器3に対するユーザの指示（例えば、歌詞の設定、音色等の設定等）を送信してもよい。

#### 【0016】

カラオケシステムでは、制御端末2は、ユーザによって設定された楽曲を設定する指示を、カラオケ装置1に送信する。カラオケ装置1は、当該指示に基づき、カラオケ用サーバ1000から当該楽曲の楽曲データおよび歌唱音合成用サーバ2000から第1歌詞デ

50

ータを取得する。カラオケ装置 1 は、電子楽器 3 に第 1 歌詞データを送信する。電子楽器 3 には、第 1 歌詞データが格納される。楽曲演奏を開始するユーザの指示によって、カラオケ装置 1 は、楽曲データを読み出して伴奏音等を出力し、電子楽器 3 は、第 1 歌詞データを読み出して、ユーザによる演奏操作に応じた合成歌唱音を出力する。

[ 電子楽器のハードウェア構成 ]

電子楽器 3 は、演奏操作部 3 2 1 ( 図 2 ) の操作に応じて、指示内容に従って合成歌唱音を示すオーディオ信号を生成する装置である。本実施形態では、電子楽器 3 は、電子鍵盤装置である。演奏操作部 3 2 1 は、複数の鍵を含む鍵盤および各鍵に対する操作 ( 以下、演奏操作という場合がある ) を検出するセンサを含む。本実施形態では、合成歌唱音は、電子楽器 3 からオーディオ信号がカラオケ装置 1 の入力端子に供給されることによってカラオケ装置 1 のスピーカから出力されてもよいし、電子楽器 3 に接続されたスピーカから出力されてもよい。

10

【 0 0 1 7 】

図 2 は、本発明の一実施形態における電子楽器 3 の構成を示すブロック図である。電子楽器 3 は、制御部 3 0 1、記憶部 3 0 3、操作部 3 0 5、表示部 3 0 7、通信部 3 0 9、インターフェイス 3 1 7、および演奏操作部 3 2 1 を含む。これらの各構成は、バスを介して接続されている。

【 0 0 1 8 】

制御部 3 0 1 は、CPU などの演算処理回路を含む。制御部 3 0 1 は、記憶部 3 0 3 に記憶されたプログラムを CPU により実行して、各種機能を電子楽器 3 において実現させる。電子楽器 3 において実現される機能は、例えば、音生成処理を実行するための音生成機能を含む。制御部 3 0 1 は、音生成機能によってオーディオ信号を生成するための DSP ( Digital Signal Processor ) を含む。記憶部 3 0 3 は、不揮発性メモリ等の記憶装置である。記憶部 3 0 3 は、上記音生成機能を実現するためのプログラムを記憶する。音生成機能については、後述する。また、記憶部 3 0 3 は、合成歌唱音を示すオーディオ信号を生成する際に用いる設定情報、合成歌唱音を生成するための音声素片等を記憶する。設定情報とは、例えば、音色、および歌唱音合成用サーバ 2 0 0 0 から受信した第 1 歌詞データなどである。

20

【 0 0 1 9 】

操作部 3 0 5 は、スイッチ、ボリュームつまみなどの装置であり、入力された操作に応じた信号を制御部 3 0 1 に出力する。表示部 3 0 7 は、液晶ディスプレイ、有機 EL ディスプレイ等の表示装置であり、制御部 3 0 1 による制御に基づいた画面が表示される。なお、操作部 3 0 5 と表示部 3 0 7 とは一体としてタッチパネルを構成してもよい。通信部 3 0 9 は、制御部 3 0 1 の制御に基づいて、短距離無線通信によって制御端末 2 と接続する。

30

【 0 0 2 0 】

演奏操作部 3 2 1 は、演奏操作に応じた演奏信号を制御部 3 0 1 に出力する。演奏信号は、操作された鍵の位置を示す情報 ( ノートナンバ )、押鍵したことを示す情報 ( ノートオン )、離鍵したことを示す情報 ( ノートオフ )、および押鍵速度 ( ベロシティ ) 等を含む。具体的には、押鍵されると、ベロシティおよびノートナンバ ( 音高指示ともいう ) が対応付けられたノートオンが発声の開始指示を示す演奏信号として出力され、離鍵するとノートナンバが対応付けられてノートオフが発声の停止指示を示す演奏信号として出力される。制御部 3 0 1 は、この演奏信号を用いてオーディオ信号を生成する。インターフェイス 3 1 7 は、生成されたオーディオ信号を出力する端子を含む。

40

【 0 0 2 1 】

ここで、記憶部 3 0 3 に記憶される第 1 歌詞データの一例について、図 3 を参照して説明する。図 3 は、本発明の一実施形態で使用される第 1 歌詞データである。第 1 歌詞データは、発声すべき歌詞 ( 文字 ) を規定するデータである。第 1 歌詞データは、発声すべき複数の文字が時系列に並べられたテキストデータを有する。第 1 歌詞データは、所定の時間軸により文字ごとに発声の開始時刻および停止時刻が規定されるタイミングデータを含

50

む。開始時刻および停止時刻は、例えば、楽曲の最初を基準とした時刻として定義される。このタイミングデータは、歌唱曲の進行位置と、その進行位置において発声すべき歌詞とを対応付けている。

【0022】

以降、発声すべき歌詞（文字）の各々、つまり音声上の一単位（一まとまりの音のくぎり）を、「音節」と表現することもある。本実施の形態では、歌詞データ（後述する第2歌詞データを含む）における「文字」は「音節」と同義に用いられる。

【0023】

図3に示すように、第1歌詞データは、「こ（k o）」「ん（n）」「に（n i）」「ち（c h i）」「わ（w a）」「さ（s a）」「よ（y o）」「お（o）」「な（n a）」「ら（r a）」を示すテキストデータを含む。「こ」「ん」「に」「ち」「わ」「さ」「よ」「お」「な」「ら」示す文字には、M（i）が対応付けられ、「i」（i = 1 ~ n）により歌詞における文字の順序が設定されている。例えば、M（5）は、歌詞のうち5番目の文字に対応する。第1歌詞データは、各文字M（i）に発声の開始時刻 $t_s(i)$ および停止時刻 $t_e(i)$ が設定されたタイミングデータを含む。例えば、M（1）「こ」の場合、発声の開始時刻は時刻 $t_s(1)$ であり、停止時刻は時刻 $t_e(1)$ である。同様に、M（n）「る」の場合、発声の開始時刻は時刻 $t_s(n)$ であり、停止時刻は時刻 $t_e(n)$ である。各文字M（i）に対応する時刻 $t_s(i)$  ~ 時刻 $t_e(i)$ の期間を、文字M（i）の発声の設定期間という。この発声の設定期間は、例えば理想的に歌唱した場合における期間を示している。以下に説明するように、合成歌唱音に含まれる各文字の発声期間は、演奏信号による発声の開始指示および発声の停止指示に基づいて制御されるものであって、タイミングデータに規定される発声の設定期間とは関連しない。

[音生成処理]

次に、本発明の一実施形態に係る音生成処理について、図4 ~ 図7を参照して説明する。音生成処理は、演奏操作部321への演奏操作に基づいて、各文字の発声に対応するオーディオ信号を生成する指示又は停止する指示を出力する。

【0024】

図4は、本発明の一実施形態における音生成処理を説明するフローチャートである。この処理は、制御部301のCPUが、記憶部303に記憶されたプログラムを記憶部303のRAM等に展開して実行することにより実現される。この処理は、例えば、ユーザが楽曲の再生を指示すると開始される。

【0025】

ユーザによる楽曲の再生指示により処理が開始されると、制御部301は、記憶部303から第1歌詞データを取得する（ステップS401）。次に、制御部301は初期化処理を実行する（ステップS402）。本実施形態において、初期化とは、制御部301がカウント値 $t_c = 0$ を設定することである。次に、制御部301は、カウント値 $t_c = t_c + 1$ に設定し、カウント値 $t_c$ をインクリメントする（ステップS403）。次に、伴奏データのうち、カウント値 $t_c$ に対応する部分のデータを読み出す（ステップS404）。

【0026】

制御部301は、伴奏データの読み出しの終了、ユーザによる楽曲演奏の停止指示の入力、又は演奏信号の受信、を検出するまでの間（ステップS405；No, ステップS406；No, ステップS407；No）、ステップS403及びステップS404の処理を繰り返しながら、上記の検出をするまで待機する。この状態を待機状態という。上述したように、カウント値 $t_c$ の初期値は0であり、楽曲の再生開始タイミングに対応する。制御部301は、カウント値 $t_c$ をインクリメントすることで、楽曲の再生開始タイミングを基準とした時刻を計測する。

【0027】

待機状態において伴奏データを最後まで読み出したことにより伴奏データの読み出しを終了した場合（ステップS405；Yes）、制御部301は音生成処理を終了する。待

10

20

30

40

50

機状態においてユーザによる楽曲演奏の停止指示が入力された場合（ステップ S 4 0 6 ; Y e s）、制御部 3 0 1 は音生成処理を終了する。

【 0 0 2 8 】

待機状態において演奏操作部 3 2 1 から演奏信号を受信した場合（ステップ S 4 0 7 ; Y e s）、制御部 3 0 1 は、DSPによりオーディオ信号を生成するための指示処理を実行する（ステップ S 5 0 0）。オーディオ信号を生成するための指示処理の詳細な説明については後述する。オーディオ信号を生成するための指示処理が終了すると、再びステップ S 4 0 3 に進み、制御部 3 0 1 は、ステップ S 4 0 3 およびステップ S 4 0 4 の処理を繰り返す待機状態となる。

【 0 0 2 9 】

図 5 は、図 4 のステップ S 5 0 0 で実行される指示処理を示すフローチャートである。

【 0 0 3 0 】

演奏操作部 3 2 1 から演奏信号を受信すると、オーディオ信号を生成するための指示処理が開始される。まず、制御部 3 0 1 は、演奏操作部 3 2 1 から取得した演奏信号に基づいて音高を設定する（ステップ S 5 0 1）。制御部 3 0 1 は、演奏操作部 3 2 1 から取得した演奏信号が、発声の開始指示か否かを判定する（ステップ S 5 0 2）。

【 0 0 3 1 】

制御部 3 0 1 は、演奏信号が発声の開始指示であると判定する場合（ステップ S 5 0 2 ; Y e s）、当該発声の開始指示を取得したときのカウンタ値  $t_c$  が、第 1 歌詞データを参照して、いずれかの文字に対応する発声の設定期間内であるか否かを判定する。

【 0 0 3 2 】

制御部 3 0 1 は、発声の開始指示を取得した時刻がいずれかの文字  $M(i)$  に対応する発声の設定期間内であると判定する場合（ステップ S 5 0 3 ; Y e s）、その発声の設定期間に該当する文字  $M(p)$  を発声すべき文字として設定する（ステップ S 5 0 4）。次に、制御部 3 0 1 は、設定した音高および文字  $M(p)$  の発声に基づくオーディオ信号を生成する指示を DSP に出力し（ステップ S 5 0 9）、指示処理を終了して図 4 に示すステップ S 4 0 3 に進む。

【 0 0 3 3 】

制御部 3 0 1 は、発声の開始指示を取得した時刻がいずれの文字に対しても発声の設定期間内ではないと判定する場合（ステップ S 5 0 3 ; N o）、制御部 3 0 1 は、発声の開始指示の時刻に対して直前の文字  $M(q)$  に対応する発声の停止時刻  $t_e(q)$  と、次の文字  $M(q+1)$  に対応する発声の開始時刻  $t_s(q+1)$  との間の中心時刻  $t_m(q)$  を計算する（ステップ S 5 0 5）。停止時刻  $t_e(q)$  を「第 1 時刻」とし、開始時刻  $t_s(q+1)$  を「第 2 時刻」としたとき、停止時刻  $t_e(q)$  と開始時刻  $t_s(q+1)$  との中心時刻を「第 3 時刻」と称する。カウンタ値  $t_c$  が、例えば、「こ」（文字  $M(1)$ ）の発声の停止時刻  $t_e(1)$  と「ん」（文字  $M(2)$ ）の発声の開始時刻  $t_s(2)$  との間の期間に含まれる場合、制御部 3 0 1 は、中心時刻  $t_m(1) = (t_e(1) + t_s(2)) / 2$  を計算する。なお、直前の発声の停止時刻  $t_e(q)$  と次の発声の開始時刻  $t_s(q+1)$  との間の中心時刻  $t_m(q)$  が予め計算されている場合は、ステップ S 5 0 5 を省略してもよい。次に、制御部 3 0 1 は、カウンタ値  $t_c$  が中心時刻  $t_m(q)$  よりも前か否かを判定する（ステップ S 5 0 6）。ここで、カウンタ値  $t_c$  が中心時刻  $t_m(q)$  よりも前か否かを判定することは、「第 1 条件」が成立するか否かを判定することの一例である。

【 0 0 3 4 】

カウンタ値  $t_c$  が中心時刻  $t_m(q)$  よりも前である場合（ステップ S 5 0 6 ; Y e s）、制御部 3 0 1 は、中心時刻  $t_m(q)$  よりも前の設定期間に対応する文字  $M(q)$  を設定する（ステップ S 5 0 7）。次に、制御部 3 0 1 は、設定した音高および文字  $M(q)$  の発声に基づくオーディオ信号を生成する指示を DSP に出力し（ステップ S 5 0 9）、指示処理を終了して図 4 に示すステップ S 4 0 3 に進む。

【 0 0 3 5 】

10

20

30

40

50

取得した開始指示が中心時刻  $t_m(q)$  の前ではない場合（ステップ S 5 0 6 ; N o）、制御部 3 0 1 は、中心時刻  $t_m(q)$  の後の設定期間に対応する文字  $M(q+1)$  を読み出す（ステップ S 5 0 8）。次に、制御部 3 0 1 は、取得した音高および文字の発声を開始する信号を出力し（ステップ S 5 0 9）、指示処理を終了して図 4 に示すステップ S 4 0 3 に進む。

【 0 0 3 6 】

演奏操作部 3 2 1 から取得した演奏信号が発声の開始指示ではない、つまり、発声の停止指示であると判定する場合（ステップ S 5 0 2 ; N o）、制御部 3 0 1 は、設定した音高および文字  $M(q)$  発声に基づいて生成したオーディオ信号の生成を停止する指示を DSP に出力して（ステップ S 5 1 0）、指示処理を終了して図 4 に示すステップ S 4 0 3

10

【 0 0 3 7 】

以上の指示処理をまとめると、以下のように言い換えられる。オーディオ信号を生成するための指示処理において、制御部 3 0 1 は、発声の開始指示が第 1 条件を満たすか否かを判定する。第 1 条件を満たす場合には、制御部 3 0 1 は、第 1 文字に対応する第 1 発声に基づくオーディオ信号を生成し、第 1 条件を満たさない場合には、第 1 文字の次の第 2 文字に対応する第 2 発声に基づくオーディオ信号を生成する。本実施形態において、第 1 条件とは、発声の開始指示を取得した時刻が、第 1 文字の停止時刻と第 2 文字の開始時刻との中心時刻より前であるという条件である。上記指示処理をさらに言い換えると、制御部 3 0 1 は、発声の開始指示の取得時刻が属する設定期間または取得時刻に最も近い設定期間を特定し、特定した設定期間に対応する文字に対応する発声に基づくオーディオ信号を生成する。

20

【 0 0 3 8 】

このように、順次処理することによって、楽曲の歌詞のうち伴奏音データの再生による伴奏音の進行に伴って特定される文字が、演奏操作に応じた音高およびタイミングで順次発声される合成歌唱音が生成される。そして、合成歌唱音を示すオーディオ信号がカラオケ装置 1 に対して出力される。

【 0 0 3 9 】

次に、図 4 および図 5 に示す音生成処理の具体例について、図 6 ~ 図 7 を参照して説明する。図 6 ~ 図 7 は、音生成処理における時刻と音高との関係を表す図である。

30

【 0 0 4 0 】

まず、発声の開始指示を取得したカウント値  $t_c$ （取得時刻）が、発声の設定期間  $t_s(1) \sim t_e(1)$  内である場合について説明する。音生成処理における待機状態において、制御部 3 0 1 が演奏操作部 3 2 1 から音高「G 4」が対応付けられた発声の開始指示を含む演奏信号を受信したことを想定する。この場合、制御部 3 0 1 は、指示処理を実行し（ステップ S 5 0 0）、演奏信号に基づいて、音高「G 4」を設定する（ステップ S 5 0 1）。制御部 3 0 1 は、演奏信号が発声の開始指示であると判定し（ステップ S 5 0 2 ; N o）、図 3 に示す第 1 歌詞データを参照して、開始指示を取得したカウント値  $t_c$  が発声の設定期間に含まれる（属する）か否かを判定する（ステップ S 5 0 3）。発声の開始指示を取得した時刻が設定期間  $t_s(1) \sim t_e(1)$  内であるため、制御部 3 0 1 は

40

【 0 0 4 1 】

続いて、音生成処理における待機状態において、演奏操作部 3 2 1 から音高「G 4」が対応付けられた発声の停止指示を含む演奏信号を受信したことを想定する。この場合、制御部 3 0 1 は、指示処理を実行し（ステップ S 5 0 0）、演奏信号に基づいて、音高「G

50

4」を設定する(ステップS501)。制御部301は、演奏信号が発声の停止指示であると判定し(ステップS502; No)、制御部301のDSPは、設定した音高「G4」による発声(文字「こ」)に基づくオーディオ信号の生成を停止する指示を出力する(ステップS510)。制御部301のDSPは、当該指示に基づいて、オーディオ信号の生成を停止する。

#### 【0042】

次に、発声の開始指示を取得したカウント値 $t_c$ が、発声の設定期間 $t_s(1) \sim t_e(1)$ と、設定期間 $t_s(2) \sim t_e(2)$ との間の期間であって、設定期間 $t_s(1) \sim t_e(1)$ に近い場合について、図6を参照して説明する。音生成処理における待機状態において、制御部301が演奏操作部321から音高「G4」が対応付けられた発声の開始指示を含む演奏信号を受信したことを想定する。この場合、制御部301は、指示処理を実行し(ステップS500)、演奏信号に基づいて、音高「G4」を設定する(ステップS501)。制御部301は、演奏信号が発声の開始指示であると判定し(ステップS502; No)、図3に示す第1歌詞データを参照して、開始指示を取得したときのカウント値 $t_c$ が発声の設定期間に含まれるか否かを判定する(ステップS503)。開始指示を取得した時刻が各文字 $M(i)$ に対応する発声の設定期間のいずれにも含まれないため、制御部301は、開始指示が発声の設定期間内に含まれないと判定する(ステップS503; No)。次に、制御部301は、カウント値 $t_c$ の直前および直後に設定された設定期間から中心時刻 $t_m(i)$ を計算する。開始指示を取得した時のカウント値 $t_c$ が設定期間 $t_s(1) \sim t_e(1)$ と設定期間 $t_s(2) \sim t_e(2)$ の間である場合、制御部301は、停止時刻 $t_e(1)$ と、開始時刻 $t_s(2)$ との間の中心時刻 $t_m(1)$ を計算する(ステップS505)。ここでは、 $t_m(1) = (t_e(1) + t_s(2)) / 2$ が得られる。次に、制御部301は、開始指示を取得した時のカウント値 $t_c$ が中心時刻 $t_m(1)$ よりも前であると判定し(ステップS506; Yes)、中心時刻 $t_m(1)$ よりも前の設定期間の文字「こ」(文字 $M(1)$ )を、発声する文字として設定する(ステップS507)。音高「G4」および文字「こ」の発声に基づくオーディオ信号の生成を開始する指示および停止する指示は、上述した方法と同様である。図6において、発声期間 $t_{on}(1) \sim t_{off}(1)$ が、音高「G4」および文字「こ」の発声に基づくオーディオ信号が生成されている期間である。

#### 【0043】

次に、発声の開始指示を取得したカウント値 $t_c$ が、発声の設定期間 $t_s(1) \sim t_e(1)$ と、設定期間 $t_s(2) \sim t_e(2)$ との間の期間であって、設定期間 $t_s(2) \sim t_e(2)$ に近い場合について、図7を参照して説明する。音生成処理を開始してから、ステップS505までの処理は、図6において説明した処理と同様であるため、説明を省略する。制御部301は、開始指示を取得した時刻が中心時刻 $t_m(1)$ よりも前ではないと判定し(ステップS506; No)、中心時刻 $t_m(1)$ よりも後の設定期間の文字「ん」(文字 $M(2)$ )を、発声する文字として設定する(ステップS508)。音高「G4」および文字「ん」の発声に基づくオーディオ信号の生成を開始する指示および停止する指示は、図6で説明した方法と同様である。図7において、期間 $t_{on}(1) \sim t_{off}(1)$ が、音高「G4」および文字「ん」に基づくオーディオ信号が生成されている期間である。

#### [音生成機能]

図8は、本発明の一実施形態における音生成機能を示す機能ブロック図である。なお、以下に説明する各機能を実現する構成の一部または全部は、ハードウェアによって実現されてもよい。

#### 【0044】

電子楽器3は、合成歌唱音を生成する音生成機能等を実現する機能ブロックとして、歌詞データ取得部31(第1の取得部)、発声制御部32(制御部)、信号生成部33および発声開始指示取得部34(第2の取得部)を含む。これらの各機能部の機能は、制御部301、記憶部303、不図示のタイマ等の協働により実現される。なお、本発明におい

て機能ブロックが信号生成部 3 3 を含むことは必須でない。

【 0 0 4 5 】

歌詞データ取得部 3 1 は、歌唱音合成用サーバ 2 0 0 0 からカラオケ装置 1 を介して曲 ID に対応した第 1 歌詞データを取得する。発声制御部 3 2 は、主として図 5 に示す指示処理を実行し、発声に基づくオーディオ信号の生成を開始する指示または停止する指示を信号生成部 3 3 に出力する。発声開始指示取得部 3 4 は、発声の開始指示を取得する。発声の開始指示は、例えば、ユーザから演奏操作部 3 2 1 を介して入力される演奏信号として取得される。

【 0 0 4 6 】

信号生成部 3 3 は、上述した DSP に対応し、発声制御部 3 2 から受信した指示に基づいて、オーディオ信号の生成を開始し、またはオーディオ信号の生成を停止する。信号生成部 3 3 によって生成されたオーディオ信号は、インターフェイス 3 1 7 を介して外部に出力される。

( 第 2 実施形態 )

本実施形態では、第 1 実施形態において説明した音生成処理とは一部異なる音生成処理について、図 4、図 9、および 図 1 0 を参照して説明する。本実施形態では、オーディオ信号を生成するための指示処理について、第 1 実施形態と異なっている。そのため、第 1 実施形態とは異なる部分を詳細に説明し、それ以外の部分は第 1 実施形態の説明を援用する。また、本実施形態では、ペロシティを、音量情報として扱うものとする。

【 0 0 4 7 】

本実施形態では、図 3 に示す第 1 歌詞データにおいて、文字  $M(i) = M(1) \sim M(10)$  を順番に発声するものとする。すなわち、第 1 歌詞データにおいて、複数の文字の発声順序が決められている。そのため、図 3 に示す第 1 歌詞データにおいて、発声の設定期間が規定されたタイミングデータは省略されてもよい。

【 0 0 4 8 】

図 4 に示すフローチャートにおいて、ユーザによる楽曲の再生指示により処理が開始されると、制御部 3 0 1 は、記憶部 3 0 3 から第 1 歌詞データを取得する (ステップ S 4 0 1)。次に、制御部 3 0 1 は初期化処理を実行する (ステップ S 4 0 2)。本実施形態において、初期化処理とは、制御部 3 0 1 が、第 1 実施形態と同様にカウント値  $t_c = 0$  を設定することである。第 2 実施形態ではさらに、初期化処理として、制御部 3 0 1 は、 $M(i)$  における文字カウント値  $i = 1$  (文字  $M(i) = M(1)$ ) を設定し、 $t_s = 0$  を設定する。「 $i$ 」は、上述したように、歌詞における文字の順序を示している。本実施形態において、 $t_s$  とは、直前の発声の開始指示を取得した時刻をいう。したがって、制御部 3 0 1 は、「 $i$ 」をインクリメントすることで、歌詞を構成する文字のうち、 $M(i)$  が示す文字を 1 つずつ進めていく。ステップ S 4 0 3 ~ S 4 0 7 における待機状態の処理は、第 1 実施形態と同様である。待機状態において、演奏操作部 3 2 1 からの演奏信号を受信した場合 (ステップ S 4 0 7 ; Yes)、オーディオ信号を生成するための指示処理を実行する (ステップ S 5 0 0)。

【 0 0 4 9 】

図 9 は、オーディオ信号を生成するための指示処理を説明するフローチャートである。この処理は、図 4 のステップ S 5 0 0 で実行される。

【 0 0 5 0 】

演奏操作部 3 2 1 から演奏信号を受信すると、オーディオ信号を生成するための指示処理が開始される。まず、制御部 3 0 1 は、演奏操作部 3 2 1 から取得した演奏信号に基づいて音高を設定する (ステップ S 5 2 1)。制御部 3 0 1 は、演奏操作部 3 2 1 から取得した演奏信号が、発声の開始指示か否かを判定する (ステップ S 5 2 2)。

【 0 0 5 1 】

制御部 3 0 1 は、演奏信号が発声の開始指示であると判定する場合 (ステップ S 5 2 2 ; Yes)、当該発声の開始指示を取得した時刻  $t_s$  が、 $t_c - t_s \leq t_{th}$  又は  $M(i) = M(1)$  のいずれかを満たすかを判定する (ステップ S 5 2 3)。ここで、 $t_c - t$

10

20

30

40

50

s は、最後に発声の開始指示を取得した時刻から現在までの経過時間である。t<sub>th</sub>は所定の期間である。制御部301は、時刻t<sub>s</sub>が、t<sub>c</sub> - t<sub>s</sub> < t<sub>th</sub>又はM(i) = M(1)のいずれかを満たす場合(ステップS523; Yes)、文字M(i)のオーディオ信号を生成する指示をDSPに出力する(ステップS526)。制御部301は、M(i) = M(1)を満たす場合、すなわち最初の発声である場合には、文字「こ」を発声すべき文字として設定し、t<sub>c</sub> - t<sub>s</sub> < t<sub>th</sub>を満たす場合は、直前の発声で設定した文字と同じ文字を発声すべき文字として設定する。次に、制御部301は、カウント値t<sub>c</sub>を、時刻t<sub>s</sub>に設定し(ステップS527)、指示処理を終了して図4に示すステップS403に進む。

#### 【0052】

制御部301は、時刻t<sub>s</sub>が、t<sub>c</sub> - t<sub>s</sub> < t<sub>th</sub>又はM(i) = M(1)のいずれも満たさない場合(ステップS523; No)、発声の開始指示で取得した音量が所定の音量よりも小さいか否かを判定する(ステップS524)。制御部301は、発声の開始指示で取得した音量が所定の音量よりも小さい場合(ステップS524; Yes)、ステップS526、S527を実行した後、指示処理を終了して図4に示すステップS403に進む。一方、制御部301は、発声の開始指示で取得した音量が所定の音量以上の場合(ステップS524; No)、文字カウント値i = i + 1を設定する(ステップS525)。そして制御部301は、文字カウント値i = i + 1で設定された文字の発声に基づくオーディオ信号を生成する指示をDSPに出力する(ステップS526)。次に制御部301は、カウント値t<sub>c</sub>を、時刻t<sub>s</sub>に設定し(ステップS527)、指示処理を終了して図4に示すステップS403に進む。

#### 【0053】

本実施形態では、第1条件とは、t<sub>c</sub> - t<sub>s</sub> < t<sub>th</sub>又はM(i) = M(1)のいずれかを満たすか否かという条件である。また、第1条件とは、t<sub>c</sub> - t<sub>s</sub> < t<sub>th</sub>又はM(i) = M(1)の双方を満たさなくても、音量が所定の音量よりも小さいという条件を満たすか否かという条件である。

#### 【0054】

このように、図4および図9に示す処理を順次処理することによって、楽曲の歌詞のうち伴奏音データの再生による伴奏音の進行に伴って特定される文字が、演奏操作に応じた音高及びタイミングで順次発声される合成歌唱音が生成される。そして、合成歌唱音を示すオーディオ信号がカラオケ装置1に対して出力される。

#### 【0055】

次に、図4および図9に示す音生成処理の具体例について、図10を参照して説明する。図10は、音生成処理における時刻と音高との関係を表す図である。図10では、音高「G4」および文字「こ」、音高「A5」および文字「ん」、音高「B5」および文字「ん」の発声が、音高情報を有する音節ノートとして例示されている。

#### 【0056】

制御部301は、音生成処理を開始すると、第1歌詞データを取得して(ステップS401)、初期化処理を実行する(ステップS402)。制御部301は、初期化処理において、文字M(i) = M(1)と、t<sub>c</sub> = 0と、t<sub>s</sub> = 0と、を設定する。音生成処理における待機状態において、制御部301が演奏操作部321から音高「G4」が対応づけられた演奏信号を受信したことを想定する(ステップS407; Yes)。この場合、制御部301は、指示処理を実行し(ステップS500)、演奏信号に基づいて、音高「G4」を設定する(ステップS521)。制御部301は、演奏信号が発声の開始指示であると判定し(ステップS522; Yes)、t<sub>c</sub> - t<sub>s</sub> < t<sub>th</sub>又はM(i) = M(1)のいずれかを満たすかを判定する(ステップS523)。制御部301は、M(i) = M(1)を満たすと判定する(ステップS523; Yes)。文字M(1)は、「こ」であるので、制御部301は、音高「G4」および文字「こ」の発声に基づくオーディオ信号を生成する指示をDSPに出力する(ステップS526)。制御部301は、カウント値t<sub>c</sub>を時刻t<sub>s</sub>として設定し(ステップS527)、指示処理を終了して図4に示すステ

10

20

30

40

50

ップS 4 0 3に進む。図10において、設定した音高「G 4」および文字「こ」の発声に基づいてオーディオ信号の生成する指示をDSPに出力した時刻 $t_s$ を、時刻 $t_{on}(1)$ と表記する。制御部301のDSPは、当該指示に基づいて、オーディオ信号の生成を開始する。

【0057】

続いて、音生成処理における待機処理において、制御部301が演奏操作部321から音高「G 4」が対応づけられた演奏信号を受信したことを想定する。この場合、制御部301は、指示処理を実行し(ステップS 5 0 0)、演奏信号に基づいて、音高「G 4」を設定する(ステップS 5 2 1)。制御部301は、演奏信号が発声の停止指示であると判定すると(ステップS 5 2 2; No)、設定した音高「G 4」および文字「こ」の発声に基づくオーディオ信号の生成を停止する指示を出力し(ステップS 5 2 8)、指示処理を終了して図4に示すステップS 4 0 3に進む。図10において、設定した音高「G 4」および文字「こ」の発声に基づいてオーディオ信号の生成を停止する指示をDSPに出力した時刻を、時刻 $t_{off}(1)$ と表記する。制御部301のDSPは、当該指示に基づいて、オーディオ信号の生成を停止する。図10において、期間 $t_{on}(1) \sim t_{off}(1)$ が、音高「G 4」および文字「こ」の発声に基づくオーディオ信号が生成されている期間である。

【0058】

続いて、音生成処理における待機処理において、制御部301が演奏操作部321から音高「A 5」が対応づけられた発声の開始指示を含む演奏信号を受信したことを想定する。この場合、制御部301は、指示処理を実行し(ステップS 5 0 0)、演奏信号に基づいて、音高「A 5」を設定する(ステップS 5 2 1)。次に、制御部301は、演奏信号が発声の開始指示であると判定し(ステップS 5 2 2; Yes)、 $t_c - t_s \leq t_{th}$ 又は $M(i) = M(1)$ のいずれかを満たすかを判定する(ステップS 5 2 3)。所定の期間 $t_{th}$ は、例えば10ms ~ 100msの範囲であり、本実施の形態では100msであるとする。 $t_c - t_s$ が100msを超えると、 $t_c - t_s \leq t_{th}$ を満たさないと判定される。ここでは、 $t_c - t_s$ は、所定の期間 $t_{th}$ よりも長いため、制御部301は、 $t_c - t_s \leq t_{th}$ および $M(i) = M(1)$ の双方を満たさないと判定し(ステップS 5 2 3; No)、音量が所定の音量よりも小さいか否かを判定する(ステップS 5 2 4)。制御部301は、音量が所定の音量以上である判定すると(ステップS 5 2 4; No)、文字カウント値 $i = i + 1$ を設定する(ステップS 5 2 5)。ここでは、文字 $M(1)$ の次の文字 $M(2)$ を設定する。文字 $M(2)$ は、「ん」であるので、制御部301は、音高「A 5」および文字「ん」の発声に基づくオーディオ信号を生成する指示をDSPに出力する(ステップS 5 2 6)。制御部301は、カウント値 $t_c$ を時刻 $t_s$ として設定し(ステップS 5 2 7)、指示処理を終了して図4に示すステップS 4 0 3に進む。図10において、期間 $t_{on}(2) \sim t_{off}(2)$ が、音高「A 5」および文字「ん」の発声に基づくオーディオ信号が生成されている期間である。

【0059】

続いて、音生成処理における待機状態において、演奏操作部321から音高「B 5」が対応付けられた発声の開始指示を含む演奏信号を受信したことを想定する。この場合、制御部301は、指示処理を実行し(ステップS 5 0 0)、演奏信号に基づいて、音高「B 5」を設定する(ステップS 5 2 1)。制御部301は、演奏信号が発声の開始指示であると判定し(ステップS 5 2 2; Yes)、 $t_c - t_s \leq t_{th}$ 又は $M(i) = M(1)$ のいずれかを満たすかを判定する(ステップS 5 2 3)。ここでは、 $t_c - t_s$ は、所定の期間 $t_{th}$ よりも短いため、 $t_c - t_s \leq t_{th}$ を満たすと判定し(ステップS 5 2 3; Yes)、音高「A 5」および文字「ん」の発声に基づくオーディオ信号を生成する指示を出力する(ステップS 5 2 6)。ここでは、実際には、制御部301は、直前の文字「ん」の発声を継続するようにオーディオ信号を生成する指示を出力する。従って、文字「ん」を継続発声するために音高「B 5」で長音である「-」の発声に基づくオーディオ信号が生成される。制御部301は、カウント値 $t_c$ を時刻 $t_s$ として設定し(ステップS

527)、指示処理を終了して図4に示すステップS403に進む。図10において、期間ton(3)~toff(3)が、音高「A5」および文字「ん」の発声に基づくオーディオ信号が生成されている期間である。

【0060】

このように、本実施形態に係る音生成処理においては、直前の発声の開始指示から、次の発声の開始指示までの期間が所定の期間よりも短い場合には、第1歌詞データの文字を進ませないようにすることができる。

【0061】

言い換えると、直前の発声の開始指示から、次の発声の開始指示までの期間が所定の期間よりも短い場合は、第2発声の開始指示が第1条件を満たす。この場合、制御部301は、第1発声の開始指示に対応する第1発声を継続するようにオーディオ信号を生成する指示を出力する。例えば、期間ton(3)~toff(3)の音節ノートには、音高「B5」で、長音である「-」が割り当てられる。

(変形例)

以上、本発明の一実施形態について説明したが、本発明の一実施形態は、以下のように様々な形態に変形することもできる。また、上述した実施形態および以下に説明する変形例は、それぞれ互いに組み合わせて適用することもできる。

【0062】

(1)先の実施形態では、1つの文字につき、1つの発声に基づくオーディオ信号を生成する場合について説明したが、本発明の一実施形態はこれに限定されない。1つのフレーズにつき、1つの発声に基づくオーディオ信号を生成する場合について、図11~図13を参照して説明する。

【0063】

ここで、記憶部303に記憶される第1歌詞データについて、図11を参照して説明する。図11は、本発明の一実施形態で使用される第1歌詞データである。図11に示す第1歌詞データは、「こ」「ん」「に」「ち」「わ」という第1フレーズ、および「さ」「よ」「お」「な」「ら」という第2フレーズを含む。「こ」「ん」「に」「ち」「わ」という第1フレーズを、1つの発声とする場合、第1発声の開始時刻はtfs(1)に対応し、停止時刻はtfe(1)に対応する。また、「さ」「よ」「お」「な」「ら」という第2フレーズを、1つの発声とする場合、第2発声の開始時刻はtfs(2)に対応し、停止時刻はtfe(2)に対応する。

【0064】

図12及び図13は、音生成処理における時刻と音高との関係を表す図である。図12及び図13は、フレーズで規定される発声期間を示している。図12及び図13では、フレーズ内における文字に対応する発声は、押鍵ごと又は第2実施形態に示す指示処理に従って進めてもよい。第1フレーズと第2フレーズとの間は、第1フレーズの停止時刻tfe(1)と、第2フレーズの開始時刻tfs(2)との間の中心時刻tfm(1)が予め設定されていてもよい。中心時刻tfm(1)は、中心時刻 $tfm(1) = (tfe(1) + tfs(2)) / 2$ を計算することで求められる。制御部301は、第1実施形態と同様に発声の開始指示の取得時刻が、中心時刻tfm(1)よりも前か否かを判定する。

【0065】

制御部301は、発声の開始指示が中心時刻tfm(1)よりも前であると判定する場合、第1フレーズの最初の文字に対応する発声に基づいてオーディオ信号を生成する指示をDSPに出力する。その後、制御部301は、発声の開始指示が中心時刻tfm(1)よりも前であると判定する場合、続けて、第2フレーズの最初の文字から当該文字に対応する発声に基づいてオーディオ信号を生成する指示をDSPに出力してもよい。

【0066】

制御部301は、発声の開始指示が中心時刻tfm(1)よりも後であると判定する場合、さらに、発声の開始指示が第2フレーズの開始時刻tfs(2)よりも後か否かを判定する。制御部301は、発声の開始指示が第2フレーズの開始時刻tfs(2)よりも

10

20

30

40

50

後であると判定すると、第2フレーズの発声に対応する文字のうち未発声の文字から当該文字に対応する発声に基づいてオーディオ信号を生成する指示をDSPに出力する。具体的には、図12に示すように、第1フレーズの開始時刻 $t_{fs}(1)$ から停止時刻 $t_{fe}(1)$ の間において、「こ」「ん」「に」「ち」「わ」「さ」の文字に対応する発声に基づいてオーディオ信号が生成されている場合を想定する。第2フレーズの開始時刻 $t_{fs}(2)$ よりも後(時刻 $t_{fon}$ )に、発声の開始指示を取得すると、第2フレーズの文字「よ」に対応する発声に基づいてオーディオ信号を生成する。時刻 $t_{foff}$ に、文字「ら」に対応する発声の停止指示を取得すると、制御部301は、オーディオ信号の生成を停止する指示をDSPに出力する。

【0067】

10

一方、制御部301は、発声の開始指示が第2フレーズの開始時刻 $t_{fs}(2)$ よりも前であると判定すると、発声に対応する文字の最初の文字に対応する発声に基づいてオーディオ信号を生成する。具体的には、図13に示すように、第1フレーズの開始時刻 $t_{fs}(1)$ から停止時刻 $t_{fe}(1)$ の間において、「こ」「ん」「に」「ち」「わ」「さ」の文字に対応する発声に基づいてオーディオ信号が生成されている場合を想定する。第2フレーズの開始時刻 $t_{fs}(2)$ よりも前(時刻 $t_{fon}$ )に、発声の開始指示を取得すると、第2フレーズの文字「さ」に対応する発声に基づいてオーディオ信号を生成する。時刻 $t_{foff}$ に、文字「ら」に対応する発声の停止指示を取得すると、制御部301は、オーディオ信号の生成を停止する指示をDSPに出力する。

【0068】

20

変形例(1)において、第1条件とは、発声の開始指示を取得した時刻が、第1フレーズの停止時刻と第2フレーズの開始時刻との中心時刻よりも前であるという条件である。また、第2条件とは、発声の開始指示を取得した時刻が、第2発声の開始時刻 $t_{fs}(2)$ よりも後であるという条件である。言い換えると、発声の開始指示の取得時刻が、第1歌詞データにおいて規定されている第2発声の開始時刻よりも後である場合に、上記第2条件は満たされる。

【0069】

(2)先の実施形態では、リードボーカルパートと第1歌詞データとが対応する合成歌唱音を生成する方法について説明したが、本発明はこれに限定されない。コーラスパートと第2歌詞データとが対応する合成歌唱音を生成する場合について、図14を参照して説明する。

30

【0070】

図14は、コーラスパートに対応する第2歌詞データである。第2歌詞データについても、発声すべき複数の文字が時系列に並べられたテキストデータを有する。第2歌詞データは、所定の時間軸により複数の文字ごとに発声の開始時刻および停止時刻が規定されるタイミングデータを含む。

【0071】

図14に示すように、第2歌詞データは、「あ(a)」「あ(a)」「あ(a)」「あ(a)」「あ(a)」「あ(a)」「お(o)」「お(o)」「お(o)」「お(o)」「お(o)」を示すテキストデータを含む。また、第2歌詞データは、文字ごとに発声の開始時刻 $t_s$ および停止時刻 $t_e$ が規定されたタイミングデータを含む。文字ごとに、 $N(i)$ が対応づけられ、「 $i$ 」( $i=1\sim n$ )により歌詞における文字の順序が設定されている。例えば、 $N(3)$ は、歌詞のうちの3番目の文字に対応する。例えば、 $N(3)$ 「あ」の場合、発声の開始時刻は時刻 $t_{cs}(3)$ であり、停止時刻は時刻 $t_{ce}(3)$ である。

40

【0072】

図3に示すように、第1歌詞データに規定される発声期間と、図14に示すように、第2歌詞データに規定される発声期間とは、重複している。すなわち、図14に示す $N(1)\sim N(n)$ における開始時刻および終了時刻と、図3に示す $M(1)\sim M(n)$ における開始時刻および終了時刻とは、互いに時間的に一致している。この場合には、制御部301は、リードボーカルパートに代えてコーラスパートの文字に対応する発声に基づくオ

50

オーディオ信号を生成する指示をDSPに出力してもよい。また、制御部301は、第1歌詞データに規定される発声期間と、第2歌詞データに規定される発声期間とが重複する場合には、第1実施形態における第1条件を、他の条件に変更してもよい。他の条件として、直前の文字M(q)に対応する発声の停止時刻 $t_e(q)$ と、次の文字M(q+1)に対応する発声の開始時刻 $t_s(q+1)$ との間の中心時刻 $t_m(q)$ を、中心ではなく前後にずらしてもよい。例えば、 $t_m(q) = (t_e(q) + t_s(q+1)) \times (1/3)$ としてもよいし、 $t_m(q) = (t_e(q) + t_s(q+1)) \times (2/3)$ としてもよい。

#### 【0073】

なお、次のように制御してもよい。制御部301は、第1歌詞データにおいて、発声の開始指示の取得時刻が属する設定期間または取得時刻に最も近い設定期間を特定する。そして制御部301は、上記特定した設定期間と時間的に一致する設定期間が第2歌詞データに存在する場合は、第1歌詞データにおける第1発声または第2発声に基づくオーディオ信号に代えて、第2歌詞データにおける上記時間的に一致する設定期間に対応する文字に対応する発声に基づくオーディオ信号を生成する。すなわち、発声の開始指示の取得時刻に対応する設定期間が第1歌詞データと第2歌詞データとの双方に存在する場合は、第2歌詞データの発声を優先させる。このような処理は、第1歌詞データに対して一部の時間領域においてだけ第2歌詞データが対応している場合にも適用できる。なお、コーラスパートを併用する場合は、上記第3時刻を、停止時刻 $t_e(q)$ と開始時刻 $t_s(q+1)$ との中心時刻に対して前または後ろにずらしてもよい。

#### 【0074】

(3)本実施形態では、電子楽器3は電子鍵盤装置である場合について説明したが、これに限定されない。電子楽器3は電子管楽器であってもよい。以下、電子楽器3として電子管楽器を適用する場合について、図15を参照して説明する。

#### 【0075】

図15は、電子楽器3Aが電子管楽器である場合のハードウェア構成である。電子管楽器の場合、演奏操作部321は、操作キー311およびプレスセンサ312を含む。

#### 【0076】

電子楽器3Aは、楽器本体に設けられた複数の音孔と、音孔の開閉状態を変化させる複数の操作キー311と、プレスセンサ312が設けられている。複数の操作キー311を演奏者が演奏することによって、音孔の開閉状態が変化して所定の音階の音を出力する。楽器本体には、マウスピースが取り付けられており、楽器本体の内部であって、マウスピースの近傍には、プレスセンサ312が設けられている。プレスセンサ312は、マウスピースを介してユーザ(演奏者)が吹き込む息の吹圧を検出する吹圧センサである。プレスセンサ312は、息の吹込みの有無を検出するとともに、少なくとも電子楽器3Aの演奏時においては、吹圧の強さや速さ(勢い)を検出する。プレスセンサ312により検出された圧力の大きさに応じて、発声の音量が決定される。本変形例では、プレスセンサ312により検出した圧力の大きさを、音量情報として扱うものとする。プレスセンサ312において、所定の圧力の大きさが検出された場合には、発声の開始指示として検出される。所定の圧力未満の場合には、発声の開始指示として検出されない。

#### 【0077】

電子管楽器において、図9および図10で説明したように、第1発声の開始指示から第2発声の開始指示までの第1期間が所定の期間未満となる、管楽器特有の経過音として検出されることがある。本発明の一実施形態に係る音生成処理では、演奏の途中でこのような経過音が生じたとしても、歌詞の位置が演奏よりも先に進むことを抑制することができるため、自然な合成歌唱音を生成することができる。

#### 【0078】

(4)第1実施形態において、中心時刻 $t_m(q) = (t_e(q) + t_s(q+1)) / 2$ とする場合について説明したが、これに限定されない。中心ではなく前後にずらしてもよい。例えば、 $t_m(q) = (t_e(q) + t_s(q+1)) \times (1/3)$ としてもよ

10

20

30

40

50

いし、 $t_m(q) = (t_e(q) + t_s(q+1)) \times (2/3)$ としてもよい。

【0079】

(5) 第2実施形態において、第1条件は、音量が所定の音量よりも小さいという条件を満たすか否かという条件を含む場合について説明したが、本発明の一実施形態はこれに限定されない。第1条件は、図9において、ステップS524を省略してもよく、ステップS523における $t_c - t_s$   $t_{th}$ 又は $M(i) = M(1)$ のいずれかを満たすか否かという条件だけであってもよい。

【0080】

以上、本発明をその好適な実施形態に基づいて詳述してきたが、本発明はこれら特定の実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の様々な形態も本発明に含まれる。上述の実施形態の一部を適宜組み合わせてもよい。

10

【0081】

なお、演奏信号は、外部から通信を介して取得されてもよい。従って、演奏操作部321を設けることは必須でなく、音生成装置は、楽器としての機能および形態を有することは必須でない。

【0082】

なお、本発明を達成するためのソフトウェアによって表される制御プログラムを記憶した記憶媒体を、本装置に読み出すことによって、本発明と同様の効果を奏するようにしてもよく、その場合、記憶媒体から読み出されたプログラムコード自体が本発明の新規な機能を実現することになり、そのプログラムコードを記憶した、非一過性のコンピュータ読み取り可能な記録媒体は本発明を構成することになる。また、プログラムコードを伝送媒体等を介して供給してもよく、その場合は、プログラムコード自体が本発明を構成することになる。なお、これらの場合の記憶媒体としては、ROMのほか、フロッピディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード等を用いることができる。非一過性のコンピュータ読み取り可能な記録媒体としては、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリ(例えばDRAM(Dynamic Random Access Memory))のように、一定時間プログラムを保持しているものも含む。

20

【0083】

本出願は、2021年3月9日に出願された日本出願第2021-037651号に基づく優先権を主張するものであり、当該日本出願に記載された全内容を本出願に援用する。

30

【符号の説明】

【0084】

### 3 電子楽器

31 歌詞データ取得部

32 発声制御部

34 発声開始指示取得部

301 制御部

40

321 演奏操作部

【手続補正2】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

発声すべき複数の文字が時系列に並べられ、第1文字および当該第1文字の後の第2文字を少なくとも含む第1歌詞データを取得する第1の取得部と、

50

発声の開始指示を取得する第2の取得部と、

前記第2の取得部により前記発声の開始指示が取得された場合において、前記発声の開始指示が第1条件を満たす場合には、前記第1歌詞データにおける前記第1文字に対応する第1発声に基づくオーディオ信号を生成する指示を出力し、前記発声の開始指示が前記第1条件を満たさない場合には、前記第1歌詞データにおける前記第2文字に対応する第2発声に基づくオーディオ信号を生成する指示を出力する制御部と、を有し、

前記第1歌詞データにおいて、所定の時間軸により、各文字の発声の開始時刻および停止時刻が規定され、

前記発声の開始指示の取得時刻が、前記所定の時間軸における、前記第1発声を停止する第1時刻と前記第2発声を開始する第2時刻との間に位置する第3時刻より前である場合に、前記発声の開始指示が前記第1条件を満たす、音生成装置。

10

【請求項2】

前記第3時刻は、前記第1時刻と前記第2時刻との中心時刻である、請求項1に記載の音生成装置。

【請求項3】

前記第1発声および前記第2発声は、一つの文字に対応する発声である、請求項1または2に記載の音生成装置。

【請求項4】

前記第1発声は、前記第1文字を含む複数の文字に対応する発声を含み、

前記第2発声は、前記第2文字を含む複数の文字に対応する発声を含み、

20

前記第1時刻は、前記第1発声に対応する前記複数の文字のうち最後の文字の発声を停止する時刻に対応し、

前記第2時刻は、前記第2発声に対応する前記複数の文字のうち最初の文字の発声を開始する時刻に対応する、請求項1または2に記載の音生成装置。

【請求項5】

前記制御部は、前記第1発声に対応する前記最後の文字の発声の後において前記第2発声の最初の文字から当該文字に対応する発声に基づいてオーディオ信号を生成する指示を出力し、

前記制御部は、前記第2発声に基づくオーディオ信号を生成する指示を出力する際、前記発声の開始指示がさらに第2条件を満たしている場合に、前記第2発声に対応する文字のうち未発声の文字から当該文字に対応する発声に基づいてオーディオ信号を生成する指示を出力する、請求項4に記載の音生成装置。

30

【請求項6】

前記発声の開始指示の取得時刻が、前記第1歌詞データにおいて規定されている前記第2発声の開始時刻よりも後である場合に、前記第2条件は満たされる、請求項5に記載の音生成装置。

【請求項7】

発声すべき複数の文字が時系列に並べられ、第1文字および当該第1文字の後の第2文字を少なくとも含む第1歌詞データを取得する第1の取得部と、

発声の開始指示を取得する第2の取得部と、

40

前記第2の取得部により前記発声の開始指示が取得された場合において、前記発声の開始指示が第1条件を満たす場合には、前記第1歌詞データにおける前記第1文字に対応する第1発声に基づくオーディオ信号を生成する指示を出力し、前記発声の開始指示が前記第1条件を満たさない場合には、前記第1歌詞データにおける前記第2文字に対応する第2発声に基づくオーディオ信号を生成する指示を出力する制御部と、を有し、

前記第1歌詞データにおける前記文字の各々には、所定の時間軸により、発音の開始時刻および停止時刻からなる設定期間が対応付けられており、

前記制御部は、前記所定の時間軸における、前記発声の開始指示の取得時刻が属する設定期間または前記取得時刻に最も近い設定期間を特定し、特定した設定期間に対応する文字に対応する発声に基づくオーディオ信号を、前記第1発声に基づくオーディオ信号また

50

は前記第 2 発声に基づくオーディオ信号として生成する指示を出力する、音生成装置。

【請求項 8】

前記第 1 の取得部は、発声すべき複数の文字が時系列に並べられ、前記所定の時間軸により、前記各文字の発音の開始時刻および停止時刻が規定された第 2 歌詞データを取得し、

前記第 2 歌詞データにおける前記文字の各々には、前記所定の時間軸により、発音の開始時刻および停止時刻からなる設定期間が対応付けられており、

前記制御部は、前記特定した設定期間と時間的に一致する設定期間が前記第 2 歌詞データに存在する場合は、前記第 1 発声または前記第 2 発声に基づくオーディオ信号に代えて、前記第 2 歌詞データにおける前記時間的に一致する設定期間に対応する文字に対応する発声に基づくオーディオ信号を生成する指示を出力する、請求項 7 に記載の音生成装置。

【請求項 9】

発声すべき複数の文字が時系列に並べられ、第 1 文字および当該第 1 文字の後の第 2 文字を少なくとも含む第 1 歌詞データを取得する第 1 の取得部と、

発声の開始指示を取得する第 2 の取得部と、

前記第 2 の取得部により前記発声の開始指示が取得された場合において、前記発声の開始指示が第 1 条件を満たす場合には、前記第 1 歌詞データにおける前記第 1 文字に対応する第 1 発声に基づくオーディオ信号を生成する指示を出力し、前記発声の開始指示が前記第 1 条件を満たさない場合には、前記第 1 歌詞データにおける前記第 2 文字に対応する第 2 発声に基づくオーディオ信号を生成する指示を出力する制御部と、を有し、

前記第 1 歌詞データにおいて、前記複数の文字の発声順序が決められており、

前記第 2 の取得部により前記第 1 発声の開始指示の後の前記第 2 発声の開始指示が取得された場合において、前記第 1 発声の開始指示から前記第 2 発声の開始指示までの第 1 期間が所定の期間未満であるときは、前記制御部は、前記第 2 発声の開始指示が前記第 1 条件を満たすと判定し且つ、前記第 1 発声の開始指示に対応する前記第 1 発声を継続するようにオーディオ信号を生成する指示を出力する、音生成装置。

【請求項 10】

前記制御部は、前記第 2 発声の開始指示の取得時刻に対応して演奏操作部から取得される音量情報が所定値未満である場合には、前記第 1 期間が前記所定の期間以上であっても、前記第 2 発声ではなく前記第 1 発声を継続するようにオーディオ信号を生成する指示を出力する、請求項 9 に記載の音生成装置。

【請求項 11】

前記演奏操作部は、圧力変化を検出するプレスセンサを含み、

前記プレスセンサに入力された圧力変化に基づいて、前記第 1 発声の開始指示および前記第 2 発声の開始指示が取得される、請求項 10 に記載の音生成装置。

【請求項 12】

前記制御部は、前記オーディオ信号を生成する際、前記オーディオ信号に含まれる発声を演奏操作部からの音高指示に応じた音高に制御する、請求項 1 乃至 9 のいずれか 1 項に記載の音生成装置。

【請求項 13】

請求項 1 乃至 12 のいずれか 1 項に記載の音生成装置と、

前記発声の開始指示をユーザが入力するための演奏操作部と、を有する、電子楽器。

【請求項 14】

コンピュータにより実現される音生成装置の制御方法であって、

発声すべき複数の文字が時系列に並べられ、第 1 文字および当該第 1 文字の後の第 2 文字を少なくとも含む第 1 歌詞データを取得し、

発声の開始指示を取得し、

前記発声の開始指示が取得された場合において、前記発声の開始指示が第 1 条件を満たす場合には、前記第 1 歌詞データにおける前記第 1 文字に対応する第 1 発声に基づくオーディオ信号を生成する指示を出力し、前記発声の開始指示が前記第 1 条件を満たさない場

10

20

30

40

50

合には、前記第 1 歌詞データにおける前記第 2 文字に対応する第 2 発声に基づくオーディオ信号を生成する指示を出力し、

前記第 1 歌詞データにおいて、所定の時間軸により、各文字の発声の開始時刻および停止時刻が規定され、

前記発声の開始指示の取得時刻が、前記所定の時間軸における、前記第 1 発声を停止する第 1 時刻と前記第 2 発声を開始する第 2 時刻との間に位置する第 3 時刻より前である場合に、前記発声の開始指示が前記第 1 条件を満たす、音生成装置の制御方法。

【請求項 15】

音生成装置の制御方法をコンピュータに実行させるプログラムであって、

前記音生成装置の制御方法は、

発声すべき複数の文字が時系列に並べられ、第 1 文字および当該第 1 文字の後の第 2 文字を少なくとも含む第 1 歌詞データを取得し、

発声の開始指示を取得し、

前記発声の開始指示が取得された場合において、前記発声の開始指示が第 1 条件を満たす場合には、前記第 1 歌詞データにおける前記第 1 文字に対応する第 1 発声に基づくオーディオ信号を生成する指示を出力し、前記発声の開始指示が前記第 1 条件を満たさない場合には、前記第 1 歌詞データにおける前記第 2 文字に対応する第 2 発声に基づくオーディオ信号を生成する指示を出力し、

前記第 1 歌詞データにおいて、所定の時間軸により、各文字の発声の開始時刻および停止時刻が規定され、

前記発声の開始指示の取得時刻が、前記所定の時間軸における、前記第 1 発声を停止する第 1 時刻と前記第 2 発声を開始する第 2 時刻との間に位置する第 3 時刻より前である場合に、前記発声の開始指示が前記第 1 条件を満たす、プログラム。

【手続補正 3】

【補正対象書類名】図面

【補正対象項目名】全図

【補正方法】変更

【補正の内容】

10

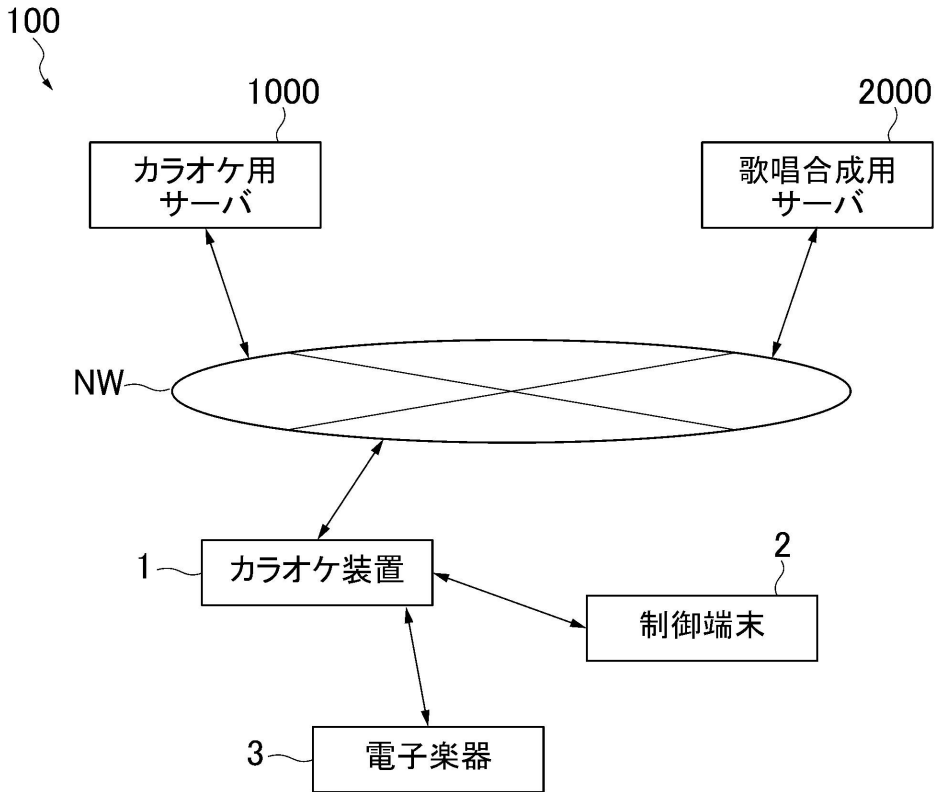
20

30

40

50

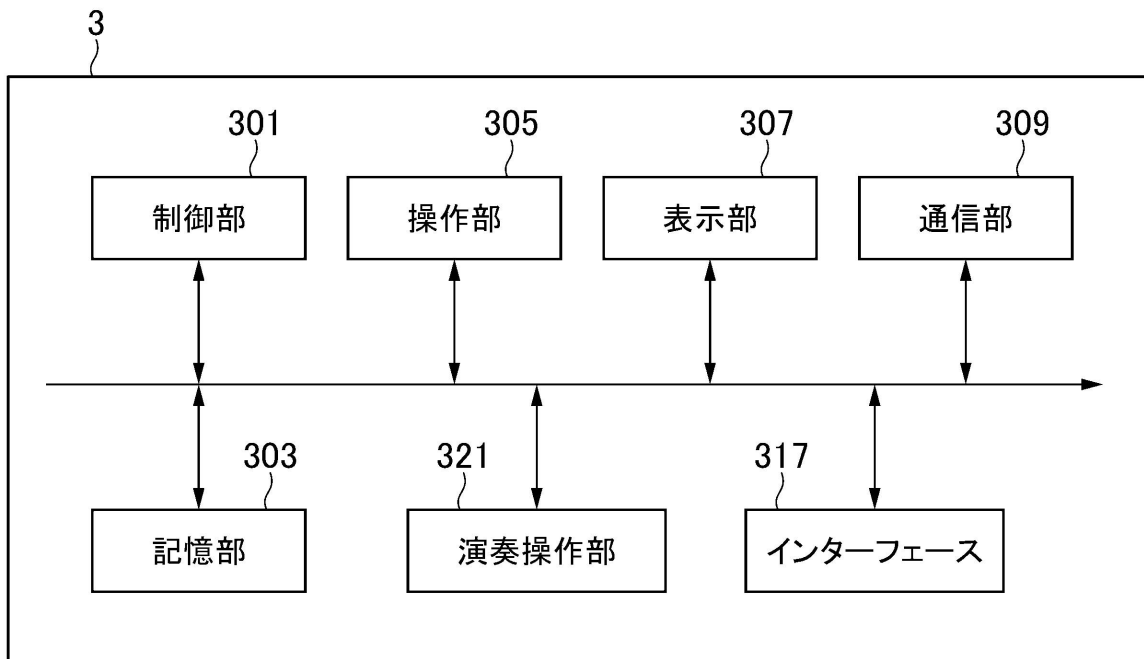
【図1】



10

20

【図2】



30

40

50

【 図 3 】

M(i)	文字	開始ts(i)	終了te(i)
M(1)	こ	ts(1)	te(1)
M(2)	ん	ts(2)	te(2)
M(3)	に	ts(3)	te(3)
M(4)	ち	ts(4)	te(4)
M(5)	わ	ts(5)	te(5)
M(6)	さ	ts(6)	te(6)
M(7)	よ	ts(7)	te(7)
M(8)	お	ts(8)	te(8)
M(9)	な	ts(9)	te(9)
M(10)	ら	ts(10)	te(10)
⋮	⋮	⋮	⋮
M(n)	る	ts(n)	te(n)

10

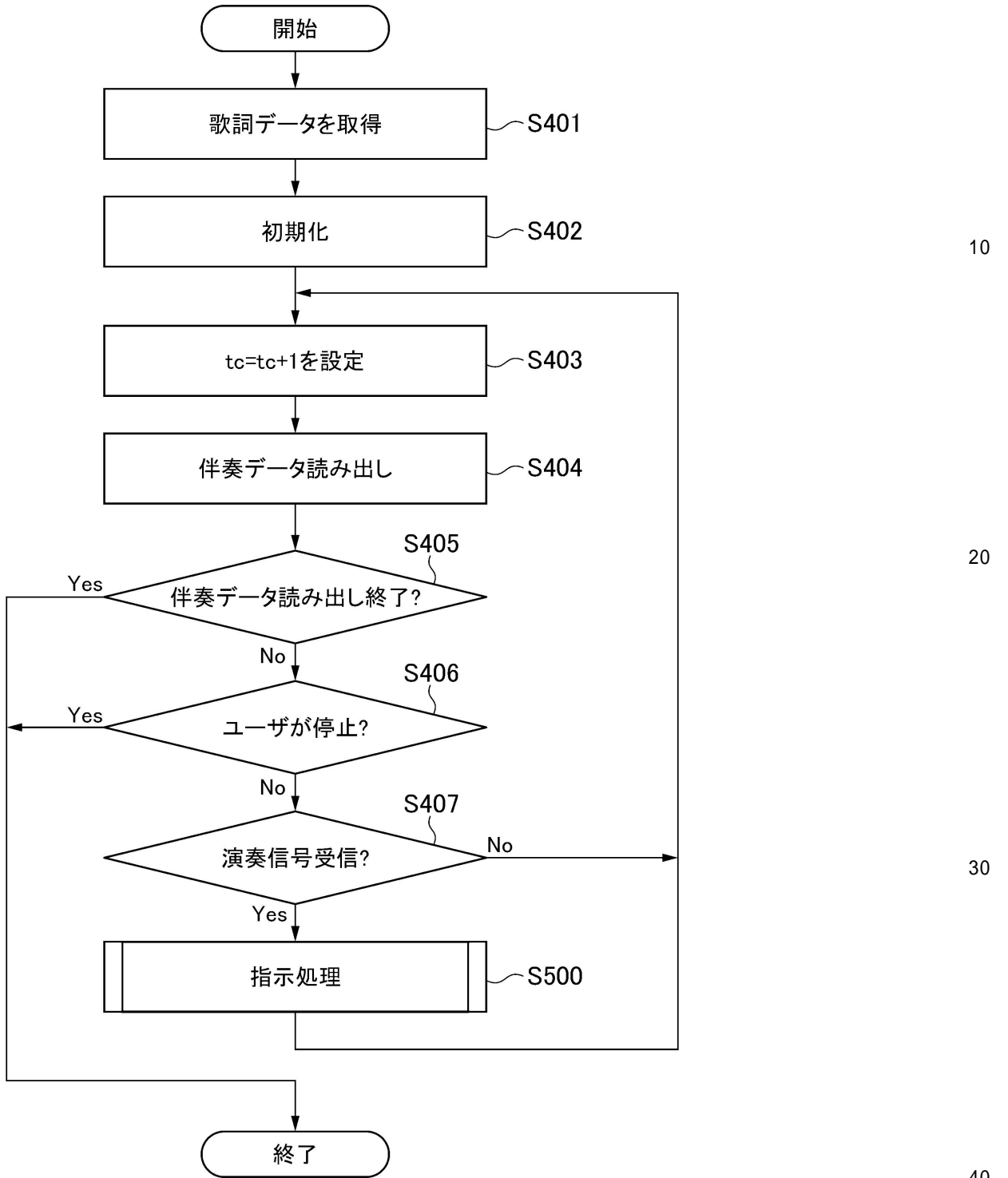
20

30

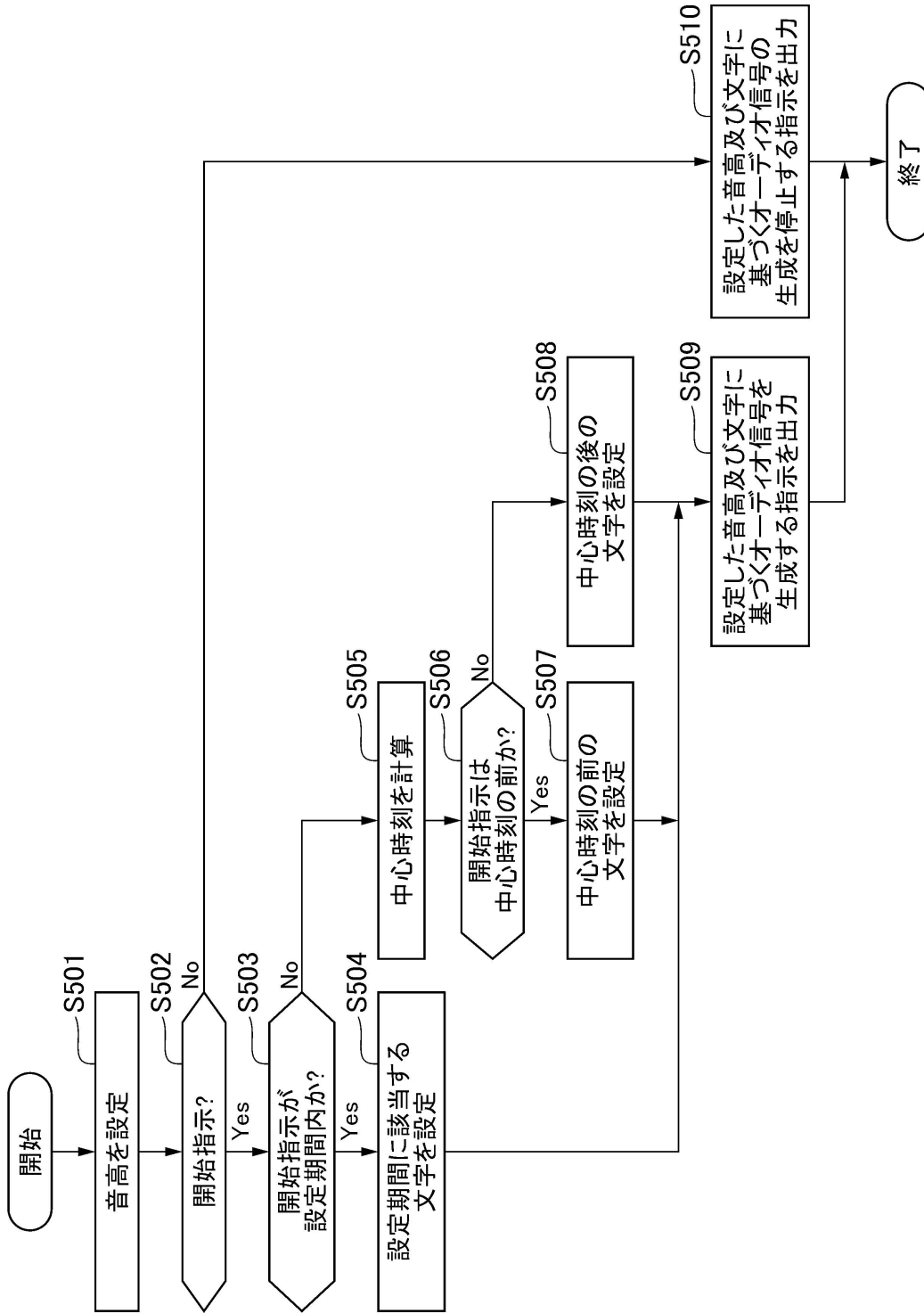
40

50

【 図 4 】



【 図 5 】



10

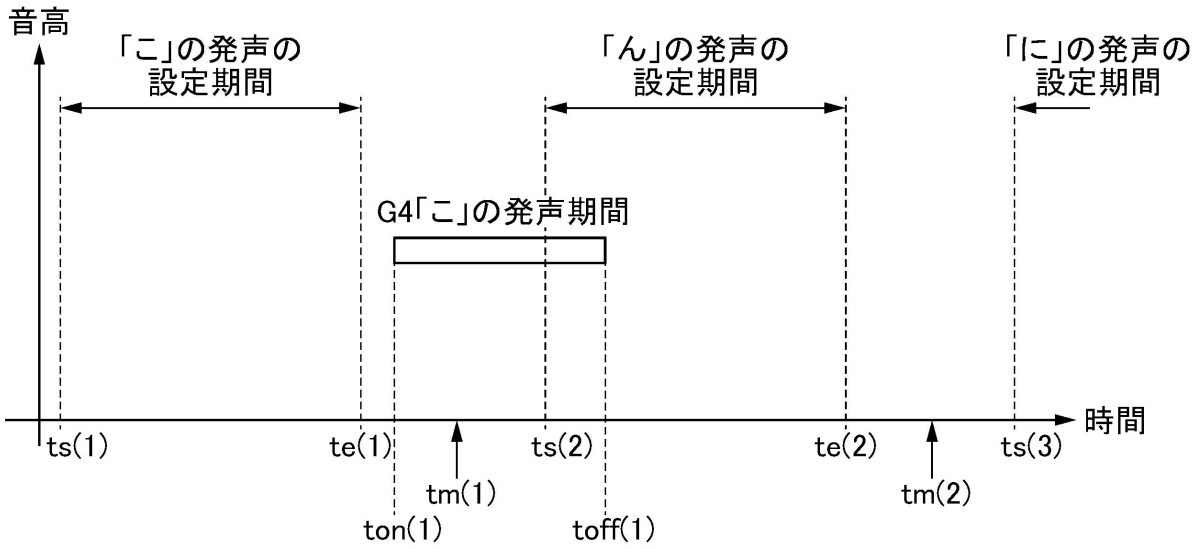
20

30

40

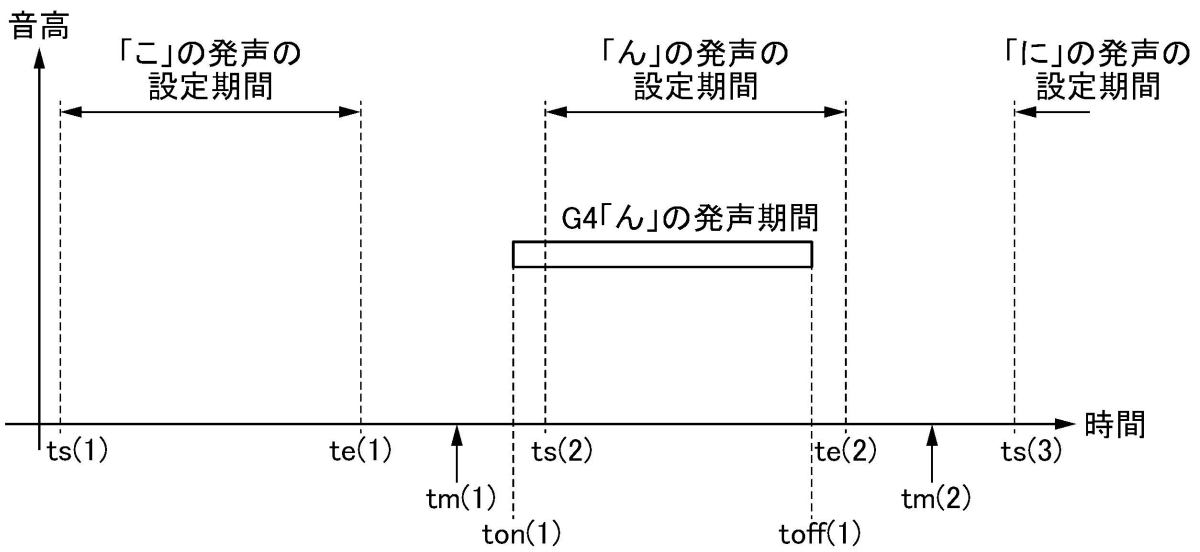
50

【図 6】



10

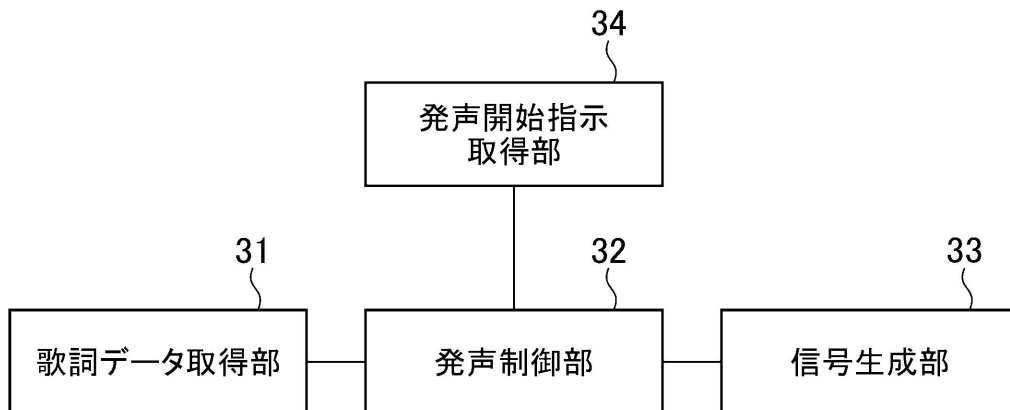
【図 7】



20

30

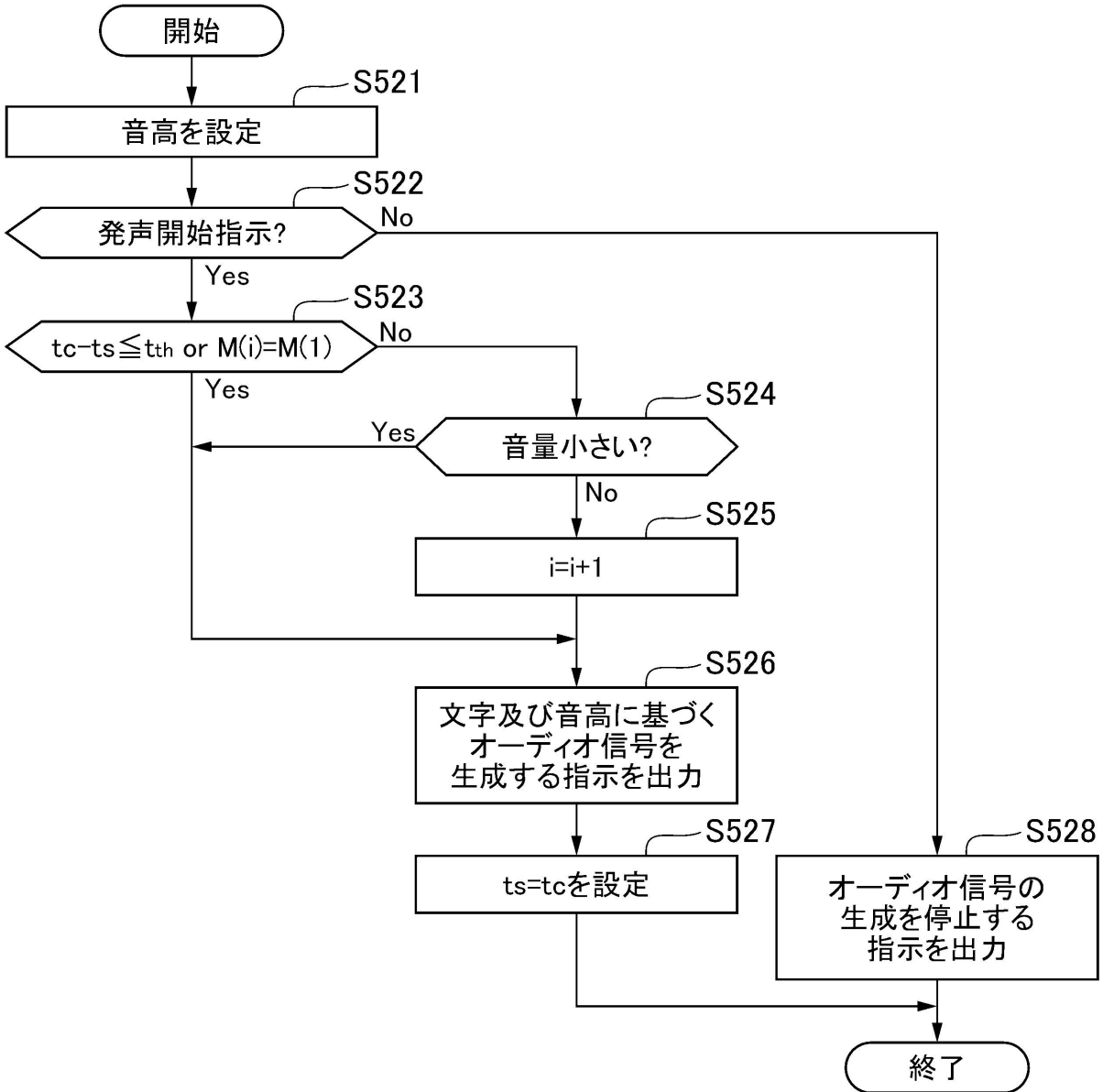
【図 8】



40

50

【図9】

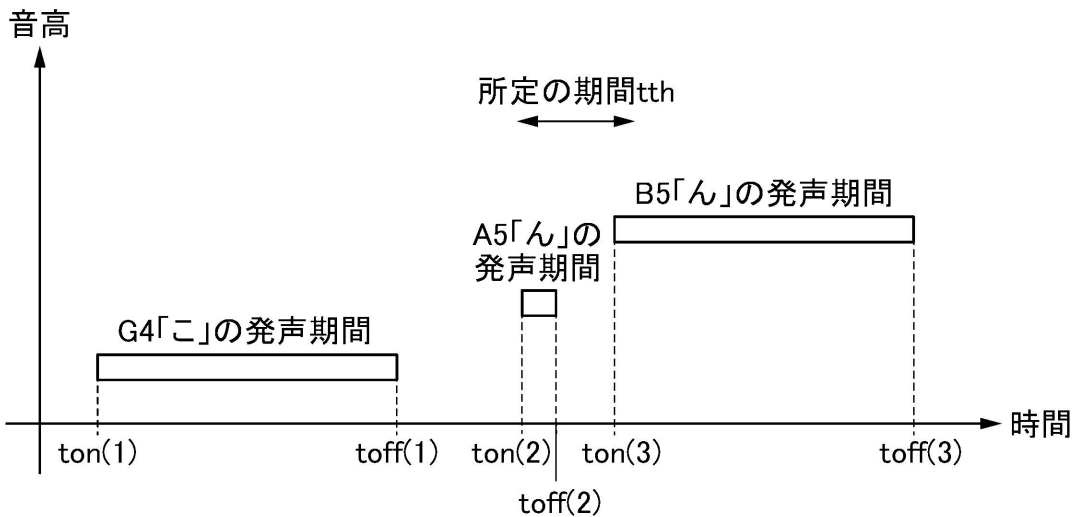


10

20

30

【図10】



40

50

【 図 1 1 】

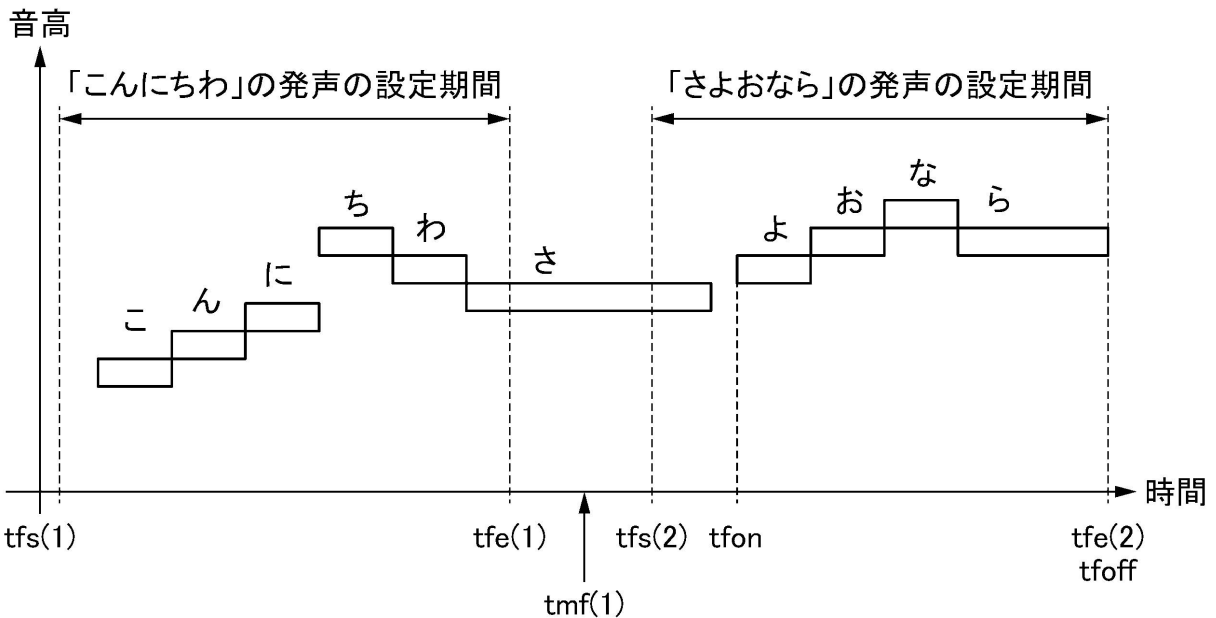
		フレーズ設定期間	
M(i)	文字	開始tfs(i)	終了tfe(i)
M(1)	こ	tfs(1)	-
M(2)	ん	-	-
M(3)	に	-	-
M(4)	ち	-	-
M(5)	わ	-	tfe(1)
M(6)	さ	tfs(2)	-
M(7)	よ	-	-
M(8)	お	-	-
M(9)	な	-	-
M(10)	ら	-	tfe(2)
⋮	⋮	⋮	⋮
M(n)	る	-	tfe(n)

10

20

30

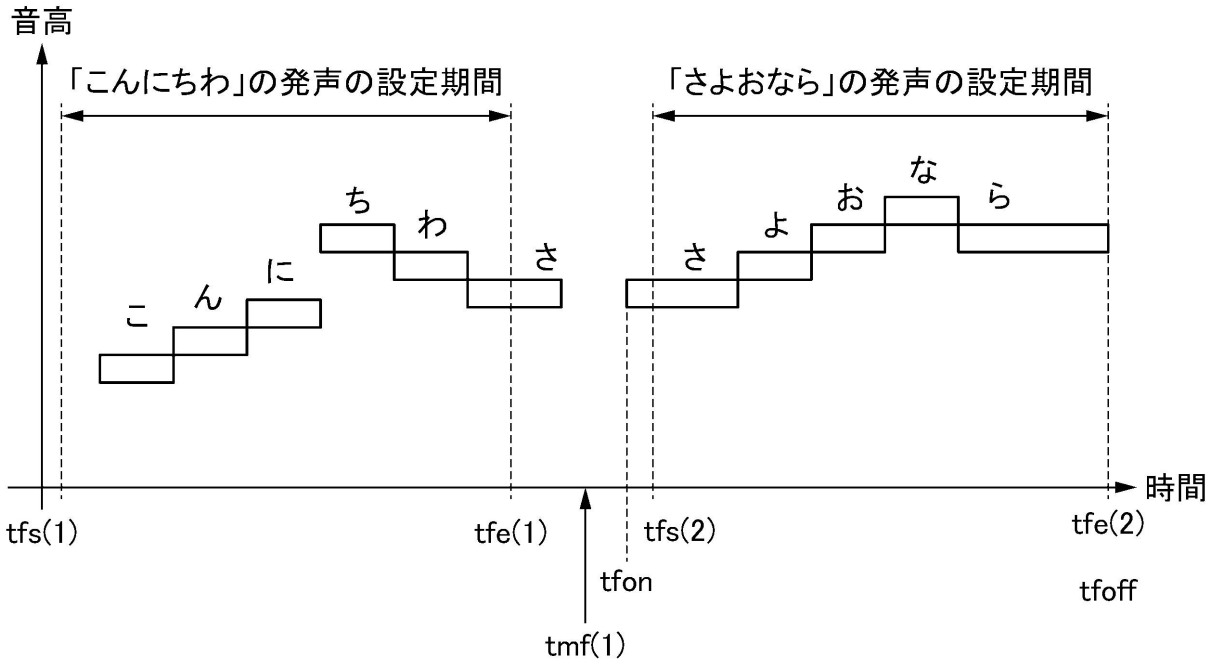
【 図 1 2 】



40

50

【 図 1 3 】



10

【 図 1 4 】

N(i)	文字	開始tcs(i)	終了tce(i)
N(1)	あ	tcs(1)	tce(1)
N(2)	あ	tcs(2)	tce(2)
N(3)	あ	tcs(3)	tce(3)
N(4)	あ	tcs(4)	tce(4)
N(5)	あ	tcs(5)	tce(5)
N(6)	お	tcs(6)	tce(6)
N(7)	お	tcs(7)	tce(7)
N(8)	お	tcs(8)	tce(8)
N(9)	お	tcs(9)	tce(9)
N(10)	お	tcs(10)	tce(10)
⋮	⋮	⋮	⋮
N(n)	る	tcs(n)	tce(n)

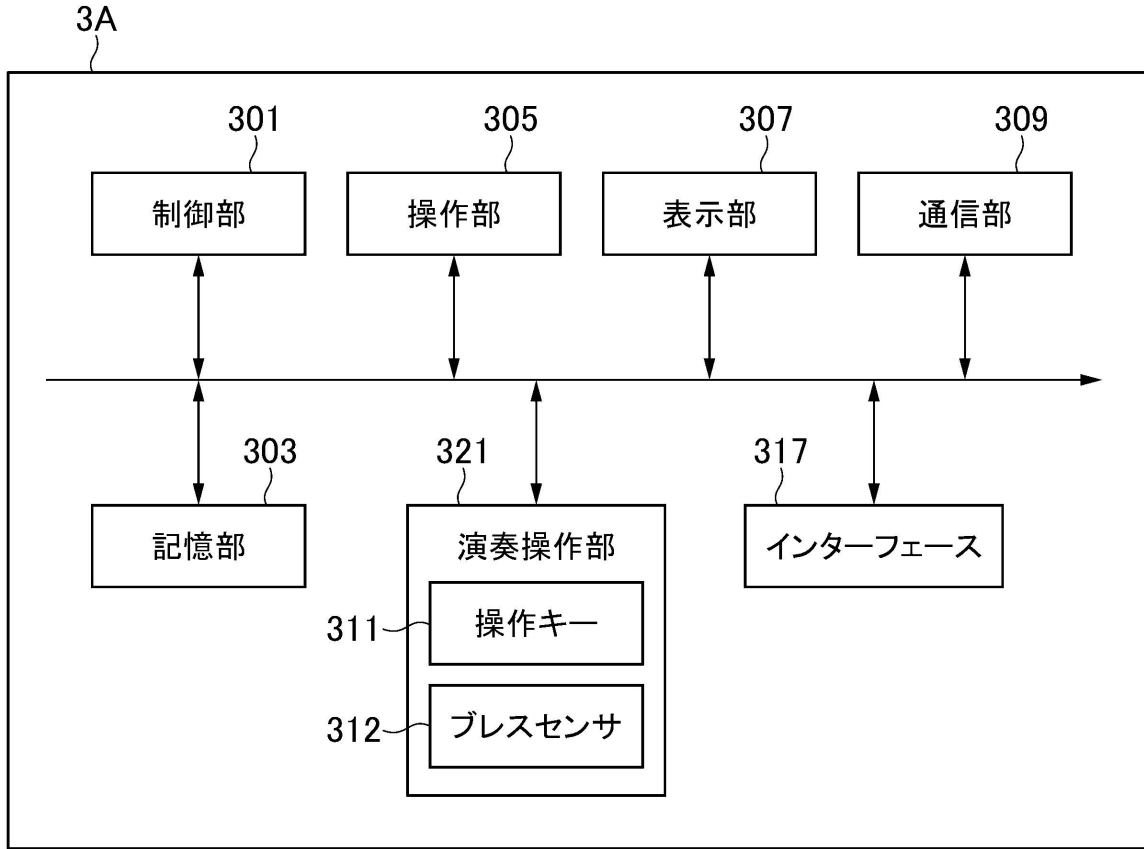
20

30

40

50

【図15】



10

20

30

40

50