

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2008年8月28日 (28.08.2008)

PCT

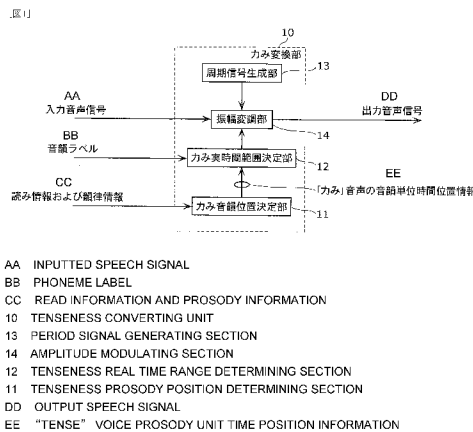
(10) 国際公開番号  
WO 2008/102594 A1

- (51) 国際特許分類:  
G10L 13/08 (2006.01) G10L 21/04 (2006.01)
- (21) 国際出願番号: PCT/JP2008/050815
- (22) 国際出願日: 2008年1月22日 (22.01.2008)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:  
特願2007-038315 2007年2月19日 (19.02.2007) JP
- (71) 出願人 (米国を除く全ての指定国について): 松下電器産業株式会社 (MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.) [JP/JP]; 〒5718501 大阪府門真市大字門真1006番地 Osaka (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 加藤 弓子 (KATO, Yumiko). 釜井 孝浩 (KAMAI, Takahiro).
- (74) 代理人: 新居 広守 (NII, Hiromori); 〒5320011 大阪府大阪市淀川区西中島5丁目3番10号タナカ・イトーピア新大阪ビル6階 新居国際特許事務所内 Osaka (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD,

[ 続葉有 ]

(54) Title: TENSENESS CONVERTING DEVICE, SPEECH CONVERTING DEVICE, SPEECH SYNTHESIZING DEVICE, SPEECH CONVERTING METHOD, SPEECH SYNTHESIZING METHOD, AND PROGRAM

(54) 発明の名称: カミ変換装置、音声変換装置、音声合成装置、音声変換方法、音声合成方法およびプログラム



(57) Abstract: A speech converting device for creating a "tense" voice appearing in a part of a speech when a person speaks tensely because of excitement, tension, anger or emphasis and more richly express a speech like speaking with anger, excitement, animation, or spirit by using variation of tone of voice. The speech converting device has a tenseness converting unit (10) comprising a tense phoneme position determining section (11) for specifying a portion to be uttered with a "tense" voice in a speech and an amplitude modulating section (14) for imparting modulation including periodical amplitude modulation to a speech waveform. The amplitude modulating section (14) imparts modulation including periodical amplitude modulation to a portion to be uttered according to the instruction of the tense phoneme position determining section (11) to create a "tense" voice and creates a speech given tensely and realistically with rich expression because of excitement, tenseness, angry, or emphasis.

(57) 要約: 興奮、緊張、怒りまたは強調のために力を入れて話すときの音声の一部に現れる「カミ」音声を生成し、怒りや興奮、勢い込んだ話し方や元気のある話し方のような音声の表情を、声質の変化により豊かに表現することができる音声変換装置に含まれるカミ変換部(10)は、音声の中の「カミ」音声で発声すべき部分を指定するカミ音韻位置決定部(11)と、周期的な振幅変動を含む変調を音声波形に付与する振幅変調部(14)とを備え、振幅変調部(14)は、カミ音韻位置決定部(11)の指定に従って「カミ」音声で発声すべき部分に周期的な振幅変動を含む変調を付与して「カミ」

[ 続葉有 ]

WO 2008/102594 A1



SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:  
— 国際調査報告書

## 明 細 書

力み変換装置、音声変換装置、音声合成装置、音声変換方法、音声合成方法およびプログラム

### 技術分野

[0001] 本発明は、(i)人間が怒鳴ったとき、強調のために力を入れて話すとき、興奮もしくは緊張して話すとき等に現れるだみ声、荒れた声、ざらざらした声 (harsh voice)、(ii)例えば演歌等の歌唱時に現れる「こぶし」や「うなり」のような表現、または(iii)ブルースやロック等の歌唱時に現れる「シャウト」のような表現といった、通常の発声とは異なる特徴を持った音声である「力み」音声を生成する技術に関する。より特定的には、本発明は、上記のような音声に含まれる(i)怒り、強勢、力強さおよび元気のよさのような感情、(ii)音声の表情、(iii)発話スタイル、または(iv)話者の態度、状況もしくは発声器官の緊張状態等を表現可能な音声の生成を可能にする音声変換装置および音声合成装置に関する。

### 背景技術

[0002] 従来、感情、表情、態度および状況等を音声で表現する、特に音声の言語的表現ではなく、口調、話し方および声色といったパラ言語的表現によって感情等を表現することを目的とした音声変換または音声合成の技術が開発されている。これらの技術は、ロボットまたは電子秘書をはじめ、電子機器の音声対話インタフェースに必要不可欠である。

[0003] 音声のパラ言語的表現のうち、韻律パターンを変更する方法については多くの方法が提案されている。モデルに基づいて基本周波数パターン、パワーパターンおよびリズムパターン等の韻律パターンを生成して、音声で表現しようとする感情に応じて、基本周波数パターンとパワーパターンとを周期変動信号により補正することにより、表現しようとする感情を伴った音声の韻律パターンを生成する方法がある(例えば、特許文献1参照)。韻律パターンの補正による感情音声の生成方法においては、特許文献1の段落0118にも指摘されているように、ゆらぎによる声質変化を防ぐために、音節の時間長を超える周期の周期変動信号が必要とされている。

- [0004] 一方、声質による表現を実現する方法としては、入力された音声进行分析して合成パラメータを求め、そのパラメータを変更して声質を変更する音声変換方法(例えば、特許文献2参照)、および標準的な音声または無表情な音声を合成するパラメータを生成し、そのパラメータを変更する音声合成方法(例えば、特許文献3参照)が開発されている。
- [0005] また、波形接続方式の音声合成技術においては、いったん標準的な音声または無表情な音声を合成し、感情等の表情のある音声の中からその合成音に類似した特徴ベクトルを持つ音声を選択して接続するものが提案されている(例えば、特許文献4参照)。
- [0006] さらに、自然音声を分析して得た合成パラメータに基づき、統計学習モデルにより合成パラメータを生成する音声合成技術においては、各感情表現を含む自然音声からそれぞれの感情に対応する音声生成モデルを統計的に学習し、モデル間の変換式を用意して、標準的な音声または無表情な音声を、感情を表現する音声に変換する方式が提案されている。
- [0007] しかしながら、上記従来の方式(方法)のうち、合成パラメータの変更を行う技術では、感情ごとにあらかじめ定められた一様な変換規則に従ってパラメータ変換を行う。このため、この技術では、自然発話に見られる部分的に力んだ声になったりするような声質のバリエーションを再現することはできない。
- [0008] また、標準的な音声と類似した特徴ベクトルを持った感情等の表情のある音声を抽出して接続する方式では、通常の発声とは大きく異なる「力み」のような特徴的な特殊な声質の音声は選択されにくい。このため、この方式では、結果的に自然発話に見られる声質のバリエーションを再現することができない。
- [0009] さらに、感情表現を含む自然音声から統計的な音声合成モデルを学習する方式では、声質のバリエーションも学習される可能性があるが、感情を表現する音声に特徴的な声質の音声は、その出現頻度が低く学習が困難である。例えば、上記の「力み」、丁寧でやさしく話すときに特徴的に現れるささやくような声(whispery voice)、およびソフトボイスともいわれるブレスリー(breathy)または「かすれ」と呼ばれる氣息性の音声(特許文献4、特許文献5参照)は、その特徴的な声質により聴取者の注意を引くた

め印象深く、発話全体の印象に大きく影響する。しかし、これらの音声は、実際の音声全体の中では一部に出現するものであり、出現頻度は高くはない。発声時間全体から見ると、その時間比率は少ないため、統計的学習を行った場合には、「力み」および「かすれ」等を再現するモデルは学習されにくい。

[0010] すなわち、上記の従来の方法では、部分的な声質のバリエーションを再現することが困難で、微細な時間構造を持った、質感のあるリアルな表情を豊かに表現することができないという課題がある。

[0011] そこで、上記の課題を解決するため、声質のバリエーションの再現のために、特徴的な声質の音声に特化した声質変換を行う方式が考えられる。声質変換の基礎となる声質の物理的特徴については、本願が対象とする「力み」とは異なる定義がされた「りきみ」音声と、上記の「かすれ」音声との研究がされている。

[0012] 「かすれ」は「息漏れ」とも呼ばれ、高調波成分でスペクトルが低く、気流による雑音成分が大きいという特徴がある。「かすれ」のこのような特徴は、「かすれ」の発声時には、通常の発声または地声(modal voice)の発声時に比べて声門の開きが大きく、「かすれ」の音声は、地声とささやき声(whisper)との中間の声であることより生じる。地声は雑音成分が少ない声であり、ささやき声は周期成分がなくなり雑音成分のみで発声される声である。「かすれ」の特徴は、第1ホルマント帯域の包絡波形と第3ホルマント帯域の包絡波形との相関の低さ、すなわち第1ホルマント付近を中心とする帯域通過信号の包絡の形状と第3ホルマント付近を中心とする帯域通過信号の包絡の形状との相関の低さとして検出される。音声合成時には上記の特徴を合成音に付加することによって「かすれ」音声を実現することができる(特許文献5参照)。

[0013] また、本願が対象とする、怒鳴ったり、興奮したりする際の発声中に生じるものとは異なる「りきみ」として、「きしる声」(creaky)または「フライ」(vocal fry)とも呼ばれる音声の研究がなされている。この研究では、「きしる声」の音響的特徴は、(i)局所的なエネルギーの変化が激しく、(ii)基本周波数は通常発声時の基本周波数よりも低く、不安定であり、(iii)通常発声の区間におけるパワーよりもパワーが小さいものとしている。これらの特徴は、発声時に喉頭を力むことにより、声帯振動の周期性が乱れるために生じる場合があることを、同研究では示している。さらに、音節単位の平均持

続時間に比べて長い区間に渡って「りきみ」が生じることが多いとしている。「きしる声」は、関心や嫌悪の感情的表現、または躊躇や謙遜のような態度的表現において、話者の誠実性を高める効果を持つ声質であるとされる。この研究で述べられる「りきみ」は、(i)一般に文末または句末等の音声が消えていく過程、(ii)言葉を選んで話したり、考えながら話したりする際に語尾を引きずるように引き伸ばして発声する場合の引き伸ばされた語尾、(iii)答えに窮した際に発せられる「えーっと」「うーん」というような感動詞または感嘆詞に多く見られるものである。さらに、この研究では「フライ」および「きしる声」にはダブル・ビートまたは基本周期の倍数で新たな周期が起きる発声(diplophonia)が含まれていることが示されている。「フライ」に見られるdiplophoniaと呼ばれる音声を生成する方式としては、基本周波数の1/2周期分位相をずらした音声を重ね合わせる方法(特許文献6参照)が提案されている。

特許文献1:特開2002-258886号公報(図8、段落0118)

特許文献2:特許第3703394号公報

特許文献3:特開平7-72900号公報

特許文献4:特開2004-279436号公報

特許文献5:特開2006-84619号公報

特許文献6:特開2006-145867号公報

特許文献7:特開平3-174597号公報

## 発明の開示

### 発明が解決しようとする課題

- [0014] しかしながら、(i)上記従来の方法では、興奮、緊張、怒りもしくは強調のために力を入れて話すときのだみ声、荒れた声もしくはざらざらした声(harsh voice)、または(ii)歌唱時の「こぶし」、「うなり」もしくは「シャウト」のような音声の一部に現れる「力み」音声を生成することはできない。ここでの「力み」音声は、力を入れて発話する際に、通常より発声器官に力が入るまたは発声器官が強く緊張するために起こるもので、発声器官が「力み」の音声を生成しやすい状況が作られる場合に発声される。具体的には、「力み」の音声は、力が入った発声であるため、音声の振幅はどちらかといえば大きく、当該モーラが両唇音または歯茎音でかつ鼻音または有声破裂音であり、文

末または句末というよりもアクセント句の先頭から3番目までの間に位置するモーラである、といった実際の音声の一部で起こる状況で発声され易い声質の音声である。また、「力み」の音声は、感動詞または感嘆詞に限らず、自立語と付属語との違いを問わず様々な品詞中に見られる。

[0015] すなわち上記の従来の方法では、本願が対象とする「力み」音声を生成することができず、怒り、興奮、緊張、勢い込んだ話し方または元気のある話し方のような音声の表情を、発声器官の力の入り方および緊張の仕方を感じさせる「力み」音声を生成することで、声質の変化により、音声を豊かに表現することが困難であるという課題を有している。

[0016] 本発明は、上記従来の課題を解決するもので、上記の「力み」音声を音声中の適切な位置に発生させることで、怒り、興奮、緊張、勢い込んだ話し方若しくは元気のある話し方、または演歌、ブルースもしくはロック等の歌唱音声において、「力み」音声を付加することにより豊かな音声表現を実現する力み音声変換装置等を提供することを目的とする。

#### 課題を解決するための手段

[0017] 本発明にある局面に係る力み変換装置は、変換の対象となる音声の部分音韻を指定する力み音韻位置指定手段と、前記力み音韻位置指定手段により指定された音韻を表す音声波形に、周期的な振幅変動を伴う変調を施す変調手段とを備える。

[0018] 後述するように、音声波形に周期的な振幅変動を伴う変調を施すことにより力み音声への変換を行なうことができる。このため、音声中の適切な音韻において力み音声を生成することができ、発声器官に力が入っている状態を、微細な時間構造を再現して、声の質感をリアルに伝える表情豊かな音声を生成することができる。

[0019] 好ましくは、前記変調手段は、前記力み音韻位置指定手段により指定された音韻を表す音声波形に、40Hz以上の周波数の周期的な振幅変動を伴う変調を施す。

[0020] さらに好ましくは、前記変調手段は、前記力み音韻位置指定手段により指定された音韻を表す音声波形に、40Hz以上でかつ120Hz以下の周波数の周期的な振幅変動を伴う変調を施す。

[0021] これにより発声器官に力が入っている状態が最も伝わり易く、かつ、人工的な歪を

感じにくい自然な音声を生成し、表情豊かな音声を生成することができる。

- [0022] 好ましくは、前記変調手段は、前記力み音韻位置指定手段により指定された音韻を表す音声波形に、振幅の変動幅を百分率で定義した周期的な振幅変動の変調度が40%以上でかつ80%以下となる、周期的な振幅変動を伴う変調を施す。
- [0023] これにより発声器官に力が入っている状態が最も伝わり易く、かつ、自然な音声を生成し、表情豊かな音声を生成することができる。
- [0024] 好ましくは、前記変調手段は、周期信号を音声波形に乗ずることにより周期的な振幅変動を伴う変調を前記音声波形に施す。
- [0025] この構成により、非常に単純な構成で力み音声を生成することができ、発声器官に力が入っている状態を、微細な時間構造を再現して、声の質感としてリアルに伝える表情豊かな音声を生成することができる。
- [0026] 好ましくは、前記変調手段は、前記力み音韻位置指定手段により指定された音韻を表す音声波形の位相をずらすオールパスフィルタと、前記力み音韻位置指定手段により指定された音韻を表す音声波形に、前記オールパスフィルタにより位相がずらされた音声波形を加算する加算手段とを備える。
- [0027] この構成により、振幅に伴って位相も変化させることができ、人工的な歪を感じにくい、より自然な変調により音声を発生させることで、感情豊かな音声を生成することができる。
- [0028] 本発明の他の局面に係る音声変換装置は、音声波形を受付ける入力手段と、力み音声に変換すべき音韻を指定する力み音韻位置指定手段と、前記力み音韻位置指定手段による力み音声に変換すべき音韻の指定に従って、前記入力手段で受け付けられた音声波形に、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を施す変調手段とを備える。
- [0029] 好ましくは、上述の音声変換装置は、さらに、前記音声波形の音韻列を認識する音韻認識手段と、前記音声波形の韻律情報を抽出する韻律分析手段とを備え、前記力み音韻位置指定手段は、前記音韻認識手段により認識された入力音声の音韻列と前記韻律分析手段により抽出された韻律情報とに基づいて力み音声に変換すべき音韻を指定する。



- [0030] この構成により、使用者は音声中の任意の音韻に力み音声を生成することができ、音声の表情を自由に表現することができる。すなわち、音声波形に周期的な振幅変動を伴う変調を施すことができ、人工的な歪を感じにくい、より自然な変調により音声を発生させることで、感情豊かな音声を生成することができる。
- [0031] 本発明のさらに他の局面に係る力み変換装置は、変換の対象となる音声中の音韻を指定する力み音韻位置指定手段と、前記力み音韻位置指定手段により指定された音韻を表す音声波形の音源信号に、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を施す変調手段とを備える。
- [0032] 音源信号に周期的な振幅変動を伴う変調を施すことにより、力み音声への変換を行なうことができる。このため、音声中の適切な音韻において力み音声を生成することができ、発声器官のうち、よりゆっくりした動きをする声道の特性を変化させず、音源波形に振幅変動を与える。これにより、発声器官に力が入っている状態を、微細な時間構造を再現して、声の質感としてリアルに伝える表情豊かな音声を生成することができる。
- [0033] なお、本発明は、このような特徴的な手段を備える力み変換装置として実現することができるだけでなく、力み変換装置に含まれる特徴的な手段をステップとする方法として実現したり、当該方法に含まれる特徴的なステップをコンピュータに実行させるプログラムとして実現したりすることもできる。そして、そのようなプログラムは、CD-ROM (Compact Disc-Read Only Memory) 等の記録媒体やインターネット等の通信ネットワークを介して流通させることができるのは言うまでもない。

### 発明の効果

- [0034] 本発明の力み変換装置等によれば、人間が怒鳴ったとき、強調のために力を入れて話すとき、および興奮または緊張して話すとき等に現れる、だみ声、荒れた声、若しくはざらざらした声 (harsh voice)、演歌等の歌唱時に現れる「こぶし」若しくは「うなり」のような表現、または、ブルース若しくはロック等の歌唱時に現れる「シャウト」のような表現といった、通常の発声とは異なる特徴を持った音声である「力み」音声を変換された音声または合成された音声の中の適切な位置に生成することができる。それにより、話者の発声器官の緊張および力の入り具合を、微細な時間構造を再現して、声

の質感としてリアルに感じさせる表情豊かな音声を生成することができる。

- [0035] また、音声波形に振幅変動を含む変調を付与する場合には簡易な処理で音声の表情を豊かにすることができる。さらに、音源波形に振幅変動を含む変調を付与する場合には、実際の「力み」音声の発声時の状態により近いと考えられる変調方式を取ることによって、人工的な歪を感じにくい、より自然な「力み」音声を生成することができる。すなわち、実際の「力み」音声において音韻性が崩れてはいないことから、「力み」の特徴は声道フィルタではなく、音源にかかわる部分で発生するものと予測される。このため、音源波形に変調を付与することが自然に起こっている現象により近い処理であると推測される。

#### 図面の簡単な説明

- [0036] [図1]図1は、本発明の実施の形態1における音声変換装置または音声合成装置に含まれる力み変換部の構成を示すブロック図である。
- [図2]図2は、実際の音声に含まれる力み音声の波形例を示す図である。
- [図3A]図3Aは、実際の音声に含まれる力みのない音声の波形と波形の包絡の概形を示す図である。
- [図3B]図3Bは、実際の音声に含まれる力みのある音声の波形と波形の包絡の概形を示す図である。
- [図4A]図4Aは、男性話者について、実際の音声に含まれる力み音声の基本周波数と振幅の変動周期との関係を示した散布図である。
- [図4B]図4Bは、女性話者について、実際の音声に含まれる力み音声の基本周波数と振幅の変動周期との関係を示した散布図である。
- [図5]図5は、実際の音声の波形とその音声に80Hzの振幅変動を施した音声の波形の図である。
- [図6]図6は、20人の被験者において、周期的な振幅変動を付加した音を「力んだ声」と判断した比率を被験者ごとに示す表である。
- [図7]図7は、聴取実験により確認された「力み」音声に聞こえる振幅変動周波数の範囲を示すグラフである。
- [図8]図8は、振幅変動の変調度を説明するための図である。

[図9]図9は、聴取実験により確認された、「カミ」音声に聞こえる振幅変動の変調度の範囲を示すグラフである。

[図10]図10は、本発明の実施の形態1における音声変換装置または音声合成装置に含まれるカミ変換部の動作を示すフローチャートである。

[図11]図11は、本発明の実施の形態1のカミ変換部の変形例の機能ブロック図である。

[図12]図12は、本発明の実施の形態1のカミ変換部の変形例の動作を示すフローチャートである。

[図13]図13は、本発明の実施の形態2における音声変換装置または音声合成装置に含まれるカミ変換部の構成を示すブロック図である。

[図14]図14は、本発明の実施の形態2における音声変換装置または音声合成装置に含まれるカミ変換部の動作を示すフローチャートである。

[図15]図15は、本発明の実施の形態2のカミ変換部の変形例の機能ブロック図である。

[図16]図16は、本発明の実施の形態2のカミ変換部の変形例の動作を示すフローチャートである。

[図17]図17は、本発明の実施の形態3における音声変換装置の構成を示すブロック図である。

[図18]図18は、本発明の実施の形態3における音声変換装置の動作を示すフローチャートである。

[図19]図19は、本発明の実施の形態3の音声変換装置の変形例の機能ブロック図である。

[図20]図20は、本発明の実施の形態3の音声変換装置の変形例の動作を示すフローチャートである。

[図21]図21は、本発明の実施の形態4における音声合成装置の構成を示すブロック図である。

[図22]図22は、本発明の実施の形態4における音声合成装置の動作を示すフローチャートである。

[図23]図23は、本発明の実施の形態4の変形例における音声合成装置の構成を示すブロック図である。

[図24]図24は、本発明の実施の形態4の変形例における入力テキストの例を示す図である。

[図25]図25は、本発明の実施の形態4の変形例における入力テキストの例を示す図である。

[図26]図26は、本発明の実施の形態4の音声合成装置の他の変形例の機能ブロック図である。

[図27]図27は、本発明の実施の形態4の音声合成装置の他の変形例の動作を示すフローチャートである。

#### 符号の説明

- [0037] 10、20 カミ変換部
- 11 カミ音韻位置決定部
- 12 カミ実時間範囲決定部
- 13 周期信号生成部
- 14 振幅変調部
- 21 オールパスフィルタ
- 22、34、45、48 スイッチ
- 23 加算器
- 31 音素認識部
- 32 韻律分析部
- 33、44 カミ範囲指定入力部
- 40 テキスト入力部
- 41 言語処理部
- 42 韻律生成部
- 43 波形生成部
- 46 カミ音韻位置指定部
- 47 切り替え入力部

## 51 力み範囲指定取得部

## 発明を実施するための最良の形態

[0038] (実施の形態1)

図1は、実施の形態1の音声変換装置または音声合成装置の一部である、力み変換部の構成を示す機能ブロック図である。図2は、「力み」音声の波形の一例を示す図である。図3Aは、実際の音声に含まれる力みのない音声の波形と波形の包絡の概形を示す図である。図3Bは、実際の音声に含まれる力みのある音声の波形と波形の包絡の概形を示す図である。図4Aは、男性話者について、実音声中に観察された「力み」音声の振幅包絡の変動周波数の分布を示した図である。図4Bは、女性話者について、実音声中に観察された「力み」音声の振幅包絡の変動周波数の分布を示した図である。図5は、通常発声の音声に「力み」の変換処理を行った音声波形の一例を示す図である。図6は、通常発声の音声と「力み」変換処理を行った音声とを聞き比べる聴取実験の結果を示したグラフである。図7は、聴取実験により確認された「力み」音声に聞こえる振幅変動周波数の範囲を示すグラフである。図8は、振幅変動の変調度を説明するための図である。図9は、聴取実験により確認された、「力み」音声に聞こえる振幅変動の変調度の範囲を示すグラフである。図10は、力み変換部の動作を示すフローチャートである。

[0039] 図1に示されるように、本発明の音声変換装置または音声合成装置の力み変換部10は、入力される音声信号を力みのある音声信号に変換する処理部であり、力み音韻位置決定部11と、力み実時間範囲決定部12と、周期信号生成部13と、振幅変調部14とを備えている。

[0040] 力み音韻位置決定部11は、音声の読み情報および韻律情報を受付け、音声の読み情報および韻律情報に基づき、対象音声の音韻ごとに力み音声で発声されるべきか否かを判断し、音韻単位で力み音声の時間位置情報を出力する処理部である。

[0041] 力み実時間範囲決定部12は、対象となる音声信号の音韻の記述を音声信号上の実時間位置に対応付けられた音韻ラベルと、上記力み音韻位置決定部11が出力する力み音声の音韻単位での時間位置情報とを受付け、音韻ラベルおよび時間位置情報に基づき、入力音声信号の実時間上での力み音声の時間範囲を決定する処理

部である。

- [0042] 周期信号生成部13は、通常の発声の音声を力み音声に変換するための周期変動信号を生成し、出力する処理部である。
- [0043] 振幅変調部14は、入力音声信号と、力み実時間範囲決定部12より出力された入力音声信号の実時間軸上での力み音声の時間範囲の情報と、周期信号生成部13より出力された周期変動信号とを受付け、入力音声信号中の指定された部分に周期変動信号を乗算することにより、力み音声を生成し、生成した力み音声を出力する処理部である。
- [0044] 実施の形態1の構成による力み変換部の動作を説明する前に、通常音声の振幅を周期的に変動させることにより、「力み」音声へ変換できることの背景について説明する。
- [0045] ここでは、本願発明に先立って、同一テキストに基づいて発話された50文について、無表情の音声と感情を伴う音声との調査を行った。感情を伴う音声のうち、「激怒」、「怒り」または「明るく元気」の感情を伴う発声において、聴取により「力み」音声であるとラベリングされた音声の多くに、図2に示すような振幅包絡が周期的に変動する波形が観察された。図2のうち、「特売してますよ」の「ばい」部分と同じ文を、感情を伴わず「平静」に発声した音声より切り出した通常発声の音声波形とその振幅包絡の概形とを図3Aに示した。また、図3Bは、図2に示した「激怒」の感情を伴って発声された「ばい」部分と同じ波形と、その振幅包絡の概形とを示す。両波形とも、音素の境界を破線で示した。図3Aの波形の“a”、“i”を発声している部分では、振幅が滑らかに変動していく様子が見える。通常の発声においては図3Aの波形のように母音の立ち上がりで滑らかに振幅が大きくなり、音素の中央付近で最大となり、音素境界に向けて小さくなる。母音の立下りがある場合には、振幅が滑らかに無音または後続子音の振幅に向けて小さくなる。図3Aのように母音が続く場合は、振幅は、緩やかに後続の母音の振幅に向けて小さくまたは大きくなる。通常発声においては、1つの母音内において、図3Bのように振幅の増減を繰り返すことはほとんどなく、このような基本周波数との関係が一見してわからない振幅の変動を持つ音声についての報告はない。そこで、本願発明者らは、「振幅変動」が「力み」音声の特徴であると考え、「力み」音

声であるとラベリングされた音声について、以下の処理によって振幅包絡の変動周期を求めた。

- [0046] まず、音声波形を代表する正弦波成分を抽出するため、対象となる音声波形の基本周波数の第2高調波を中心周波数とするバンドパスフィルタを逐次に求め、そのフィルタに音声波形を通過させる。フィルタを通過した音声波形に対してヒルベルト変換を施して解析信号を求め、その絶対値によってヒルベルト包絡曲線を求めることにより、音声波形の振幅包絡曲線を求める。求められた振幅包絡曲線をさらにヒルベルト変換し、瞬時角速度をサンプル点ごとに計算し、サンプリング周期に基づいて角速度を周波数に変換する。サンプル点ごとに求められた瞬時周波数について音韻ごとにヒストグラムを作成し、最頻値をその音韻の音声波形の振幅包絡の変動周波数と見なした。
- [0047] 図4Aおよび図4Bは、男性話者および女性話者のそれぞれについて、このような方法で求められた「カミ」音声の音韻ごとの振幅包絡の変動周波数を、音韻ごとの平均基本周波数に対してプロットした図である。男性話者、女性話者共に基本周波数に関わらず、振幅包絡の変動周波数は80Hz-90Hzを中心として、40Hz-120Hzに分布している。「カミ」音声の特徴の1つとして、40Hz-120Hzの周波数帯域に振幅の周期変動があることが発見された。
- [0048] そこで、図5に波形の例を示すような、通常発声の音声に80Hzの振幅変動を伴う変調処理を行い、図5(b)に示すような波形の処理音声、図5(a)に示すような波形の未処理音声に比べて力んで聞こえるかどうかの聴取実験を行った。6つの処理音と未処理音との組について2回ずつ聞き比べる聴取実験を20名の被験者で行ったところ、図6に示すような結果となった。80Hzの振幅変動を伴う変調処理を行った音声のほうが力んで聞こえると判断された率は、平均して82%であり、最小で42%、最大で100%、標準偏差は18%であった。この結果により、80Hzの振幅変動を伴う変調処理により通常の音声を「カミ」音声に変換することができることが確認された。
- [0049] さらに、「カミ」音声に聞こえる振幅変動周波数の範囲を確認する聴取実験をおこなった。3つの通常発声の音声に振幅変動無しから200Hzまでの15段階で振幅周波数を変えた振幅変動を伴う変調処理を行った音声を用意し、それぞれの音声がい

下の3つの分類のうちどれに該当するかを選択する実験を行った。つまり、13名の聴力正常な被験者は、通常の音声に聞こえる場合には「力みに聞こえない」を選択し、「力み」音声に聞こえる場合には「力みに聞こえる」を選択し、振幅変動が音声とは別の音を感じさせ、「力んだ声」とは聞こえない場合には「雑音に聞こえる」を選択することとした。各音声の判断は2回ずつ行われた。その結果、図7に示すように、振幅変動なしから振幅変動周波数30Hzまでは「力みに聞こえない」の回答が最も多く、振幅変動周波数40Hzから120Hzまでは「力みに聞こえる」の回答が最も多く、さらに振幅周波数130Hz以上では「雑音に聞こえる」の回答が最も多かった。この結果より、「力み」音声と判断され易い振幅変動周波数の範囲は、実際の「力み」音声の振幅変動周波数の分布に近い40Hzから120Hzであることが示された。

[0050] 一方、振幅変動の変調度は、音声波形が音韻ごとの緩やかな振幅の変動を持つため、振幅一定のキャリア信号の振幅を変調するいわゆる振幅変調とは異なる。しかし、ここでは、振幅一定のキャリア信号に対する振幅変調に習って、図8のような変調信号を仮定する。変調の対象となる信号の振幅絶対値を、1.0倍すなわち振幅の変更無しから、0倍すなわち振幅0までの間で変調する場合を変調度100%として、変調信号の変動幅を百分率で表現したものを変調度とする。図8に示した変調信号は、変調の対象の信号を変更無し(1.0倍)から、0.4倍までの間で変調するものであり、変動幅は1.0-0.4すなわち0.6である。よって変調度は60%となる。さらに、「力み」音声に聞こえる変調度の範囲を確認する聴取実験をも行った。2つの通常発声の音声に変調度0%すなわち振幅変動無しから変調度100%までの間で変調度を変えた振幅変動を伴う変調処理を行った音声を12段階用意した。これらの音声資料を聴力正常な被験者15名に聞かせ、通常の音声に聞こえる場合には“「力み」なし”、力みに聞こえる場合には“「力み」あり”、力み以外の違和感のある音に聞こえる場合には“「力み」に聞こえない”の3つの分類の中から当てはまるものを選択させる聴取実験を行った。各音声の判断は5回ずつ行われた。図9に示すように、聴取実験の結果、変調度0%から35%までは“「力み」なし”の回答が最も多く、40%から80%までは“「力み」あり”の回答が最も多かった。さらに90%以上では力み以外の違和感のある音に聞こえる、すなわち“「力み」に聞こえない”との回答が最も多かった。この



結果より、「力み」音声と判断され易い変調度の範囲は40%から80%であることが示された。

[0051] 次に、先に述べたような構成の力み変換部10の動作を図10に従って説明する。まず、力み変換部10は音声信号、音韻ラベル、ならびに音声の読み情報および韻律情報を取得する(ステップS1)。「音韻ラベル」は音韻の記述を音声信号上の実時間位置と対応付けた情報であり、「読み情報」は対象音声の発声内容を音韻列として記述したものである。「韻律情報」はアクセント句、フレーズおよびポーズといった記述的韻律情報並びに基本周波数、振幅、パワーおよび時間長といった記述的韻律情報を、音声信号として表現する際の物理量を記述した情報の少なくとも一部を含むものである。このとき、音声信号は振幅変調部14へ入力され、音韻ラベルは力み実時間範囲決定部12へ入力され、音声の読み情報と韻律情報とは力み音韻位置決定部11へ入力される。

[0052] 次に、力み音韻位置決定部11は、読み情報と韻律情報とを力み易さ推定規則に当てはめて当該音韻の力み易さを求め、力み易さがあらかじめ定められた閾値を越えた場合に、当該音韻を力み位置と決定する(ステップS2)。ステップS2で用いる推定規則は、例えば力んだ音声を含む音声データベースを用いて、あらかじめ統計的学習によって生成された推定式である。本願発明者は、このような推定規則を特許文献:国際公開第2006/123539号パンフレットに開示している。統計手法の例としては、数量化II類により、当該音韻の音韻種類、直前の音韻種類、直後の音韻種類、アクセント核からの距離、およびアクセント句内での位置といった情報を独立変数とし、当該音韻が力んだ声で発声されたか否かを従属変数として推定式を学習するものがある。

[0053] 力み実時間範囲決定部12は、力み音韻位置決定部11で音韻単位で決定された力み位置と音韻ラベルとの対応をとり、音韻単位の力み音声の時間位置情報を音声信号上の時間範囲として特定する(ステップS3)。

[0054] 一方、周期信号生成部13は、80Hzの正弦波を生成し(ステップS4)、その正弦波信号に直流成分を加えた信号を生成する(ステップS5)。

[0055] 振幅変調部14は、「力み位置」として特定された音声信号の実時間範囲について

、周期信号生成部13が生成した80Hzで振動する周期信号を入力音声信号に乗ずることで振幅変調を行い(ステップS6)、音韻の時間長より周期が短い振幅の周期的変動を含む「力み」音声への変換を行う。

[0056] かかる構成によれば、音韻ごとの情報から推定規則に従ってその音韻を力み位置とするか否かを決定し、力み位置と推定された音韻についてのみ、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を行って、適切な位置に「力み」音声を発生させる。このことにより、発声器官の緊張の度合いを感じ取ることのできる、怒り、興奮や緊張、勢い込んだ話し方、または元気のある話し方のような、微細な時間構造を持った、質感のあるリアルな感情音声を生成することができる。

[0057] なお、本実施の形態ではステップS4において周期信号生成部13は80Hzの正弦波を出力するものとしたが周波数は振幅包絡の変動周波数の分布に従い40Hz－120Hzの間のいずれの周波数でもよく、正弦波以外の周期性信号でも良い。

[0058] (実施の形態1の変形例)

図11は実施の形態1の力み変換部の変形例の機能ブロック図であり、図12は実施の形態1の力み変換部の変形例の動作を示すフローチャートである。図1および図6と同じ構成要素については同じ符号を用い、その詳細な説明は繰り返さない。

[0059] 図11に示されるように、本変形例の力み変換部10の構成は実施の形態1の図1に示す力み変換部10と同じであるが、実施の形態1において音声信号を入力として受け付けていたものを、音源波形を入力として受付けるものとする。この変更に伴い、音声波形を生成するために音源波形によって駆動される声道フィルタ61が設けられている。

[0060] 上記のような構成の力み変換部10および声道フィルタ61の動作を図12に従って説明する。まず、力み変換部10は、音源波形、音韻ラベル、ならびに音声の読み情報および韻律情報を取得する(ステップS61)。このとき、音源波形は振幅変調部14へ入力され、音韻ラベルは力み実時間範囲決定部12へ入力され、音声の読み情報と韻律情報は力み音韻位置決定部11へ入力され、声道フィルタ制御情報は声道フィルタ61に入力される。次に、力み音韻位置決定部11は、読み情報と韻律情報とを力み易さ推定規則に当てはめて、当該音韻の力み易さを求める。力み音韻位置決

定部11は、力み易さがあらかじめ定められた閾値を越えた場合に、当該音韻を力み位置と決定する(ステップS2)。力み実時間範囲決定部12は、力み音韻位置決定部11で音韻単位で決定された力み位置と音韻ラベルとの対応をとり、音韻単位の力み音声の時間位置情報を、音源波形上の時間範囲として特定する(ステップS63)。一方、周期信号生成部13は、80Hzの正弦波を生成し(ステップS4)、その正弦波信号に直流成分を加えた信号を生成する(ステップS5)。振幅変調部14は、「力み位置」として特定された音源波形の実時間範囲について、周期信号生成部13が生成した80Hzで振動する周期信号を音源波形に乗ずることによって、振幅変調を行う(ステップS66)。声道フィルタ61は、力み変換部10に入力された音源波形に対応する声道フィルタを制御するための情報(例えば、分析フレームごとのメルケプストラム係数列、または単位時間ごとのフィルタの中心周波数および帯域幅等)を入力として受け、振幅変調部14から出力される音源波形に対応する声道フィルタを形成する。振幅変調部14から出力された音源波形は、声道フィルタ61を通過し、音声波形が生成される(ステップS67)。

[0061] かかる構成によれば、実施の形態1と同様、適切な位置に「力み」音声を発生させることで、発声器官の緊張の度合いを感じ取ることのできる、怒り、興奮、緊張、勢い込んだ話し方、または元気のある話し方のような、微細な時間構造を持った、質感のあるリアルな感情音声を生成することができる。さらに、実際の「力み」音声の発声時に口や舌の振動が観察されず、音韻性も大きく損なわれてはいないことから、振幅変動は、音源または音源に近い部分で起こっていることが予測される。このため、主に口および舌の形状に関係する声道フィルタでは無く、音源波形に変調を掛けることによって、より実際の発声時の現象に近く、人工的な歪を感じにくいより自然な「力み」音声を生成することができる。ここで、音韻性とは、各音韻に特徴的に見られるスペクトル構造とその時間遷移パターンとに代表される種々の音響的特徴が見られる状態を指し、音韻性が崩れるとは音韻ごとの音響的特徴が失われ、音韻が判別可能な範囲から逸脱する状態を指す。

[0062] なお、実施の形態1と同様に、ステップS4において周期信号生成部13は80Hzの正弦波を出力するものとしたが、周波数は振幅包絡の変動周波数の分布に従い40

Hz-120Hzの間のいずれの周波数でもよく、周期信号生成部13が出力する信号は、正弦波以外の周期性信号でも良い。

[0063] (実施の形態2)

図13は、実施の形態2の音声変換装置または音声合成装置の一部である、力み変換部の構成を示す機能ブロック図である。図14は本実施の形態の力み変換部の動作を示すフローチャートである。図1および図10と同じ構成要素については同じ符号を用い、その詳細な説明は繰り返さない。

[0064] 図13に示されるように、本発明の音声変換装置または音声合成装置の力み変換部20は、入力される音声信号を力みのある音声信号に変換する処理部であり、力み音韻位置決定部11と、力み実時間範囲決定部12と、周期信号生成部13と、オールパスフィルタ21と、スイッチ22と、加算器23とを備えている。

[0065] 力み音韻位置決定部11および力み実時間範囲決定部12は、図1と同様であるので、その詳細な説明は繰り返さない。

[0066] 周期信号生成部13は、周期変動信号を生成する処理部である。

[0067] オールパスフィルタ21は、振幅応答は一定であるが位相応答が周波数によって異なるフィルタである。オールパスフィルタは電気通信の分野では伝送路の遅延特性を補償するために用いられるものであり、電子楽器の分野ではフェーザーまたはフェーズシフターと呼ばれている(非特許文献:カーティス・ロード著、青柳龍也他訳・監修「コンピュータ音楽—歴史・テクノロジー・アート—」東京電機大学出版局、p353)エフェクタ(音色に変化および効果を付加する装置)に用いられるものである。実施の形態2のオールパスフィルタ21は、位相のシフト量が可変であるという特性を持つものである。

[0068] スイッチ22は、力み実時間範囲決定部12からの入力に従って、オールパスフィルタ21の出力を加算器23へ入力するか否かを切り替えるスイッチである。

[0069] 加算器23は、オールパスフィルタ21の出力信号と入力音声信号とを加算する処理部である。

[0070] 次に、上記のような構成の力み変換部20の動作を図14に従って説明する。

[0071] まず、力み変換部20は、音声信号、音韻ラベル、ならびに音声の読み情報および

韻律情報を取得する(ステップS1)。このとき、音韻ラベルは力み実時間範囲決定部12へ入力され、音声の読み情報および韻律情報は、力み音韻位置決定部11へ入力される。また、音声信号は、加算器23へ入力される。

[0072] 次に、力み音韻位置決定部11は、実施の形態1と同様に読み情報と韻律情報とを力み易さ推定規則に当てはめて当該音韻の力み易さを求め、力み易さがあらかじめ定められた閾値を越えた場合に当該音韻を力み位置と決定する(ステップS2)。

[0073] 力み実時間範囲決定部12は、力み音韻位置決定部11で音韻単位で決定された力み位置と、音韻ラベルとの対応をとり、音韻単位の力み音声の時間位置情報を音声信号上の時間範囲として特定し(ステップS3)、スイッチ22へ切り替え信号を出力する。

[0074] 一方、周期信号生成部13は、80Hzの正弦波を生成し(ステップS4)、オールパスフィルタ21へ出力する。

[0075] オールパスフィルタ21は、周期信号生成部13より出力された80Hzの正弦波に従って位相シフト量を制御する(ステップS25)。

[0076] 入力される音声信号が力み実時間範囲決定部12より出力された「力み音声」で発声されるべき時間範囲に含まれる場合には(ステップS26でyes)、スイッチ22は、オールパスフィルタ21と加算器23とを接続し(ステップS27)、加算器23は、入力音声信号にオールパスフィルタ21の出力を加算する(ステップS28)。オールパスフィルタ21より出力された音声信号は位相シフトされているため、位相が逆転している高調波成分は、変形されていない入力音声信号と打ち消しあう。オールパスフィルタ21は、周期信号生成部13より出力された80Hzの正弦波信号に従って位相のシフト量を周期的に変動させている。このため、オールパスフィルタ21の出力と入力音声信号とを加算することで、信号が打ち消しあう量を80Hzで周期的に変動させることになる。これにより、加算結果の信号は80Hzで周期的に振幅が変動することになる。

[0077] 一方、音声信号が力み実時間範囲決定部12より出力された「力み音声」で発声されるべき時間範囲に含まれない場合には(ステップS26でno)、スイッチ22はオールパスフィルタ21と加算器23との接続を遮断し、力み変換部20は入力音声信号をそのまま出力する(ステップS29)。

- [0078] かかる構成によれば、音韻ごとの情報から推定規則に従ってその音韻を力み位置とするか否かを決定し、力み位置と推定された音韻についてのみ、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を行って、適切な位置に「力み」音声を発生させる。このことにより、発声器官の緊張の度合いを感じ取ることのできる、怒り、興奮、緊張、勢い込んだ話し方または元気のある話し方のような、微細な時間構造を持った、質感のあるリアルな感情音声を生成することができる。本実施の形態では、音韻の時間長より周期が短い周期的な振幅の変動を生成するため、すなわち音声信号のエネルギーを強めたり弱めたりするために、オールパスフィルタによって位相シフト量を周期的に変動させた信号を、原波形に加算するという方式を採用した。オールパスフィルタによる位相変化は、周波数に対して一様でない。このため、音声に含まれる様々な周波数成分には強められるものと弱められるものが混在する。実施の形態1では全ての周波数成分が同一の振幅変化をするのに対し、本実施の形態を取ることによってより複雑な振幅変化を生み出すことができ、聴感上の自然さを損ねず、人工的な歪を感じにくいという利点がある。
- [0079] なお、本実施の形態ではステップS4において周期信号生成部13が80Hzの正弦波を出力するものとしたが、周波数は40Hz-120Hzの間のいずれの周波数でもよく、正弦波以外の周期性信号でも良い。したがって、オールパスフィルタ21の位相のシフト量の変動周波数は40Hz-120Hzの間のいずれの周波数でもよく、オールパスフィルタ21は、正弦波以外の変動の特性を持つものとしても良い。
- [0080] なお、本実施の形態ではスイッチ22はオールパスフィルタ21と加算器23との接続を切り替えるものとしたが、オールパスフィルタ21への入力が入り、切りを切り替えるものとしても良い。
- [0081] なお、本実施の形態ではスイッチ22によってオールパスフィルタ21と加算器23との接続を切り替えることで力み音声変換部分と非変換部分とを切り替えたが、加算器23において、入力音声信号とオールパスフィルタ21の出力とに重みをつけて加算することにより、力み音声変換部分と非変換部分とを切り替えるものとしても良い。またはオールパスフィルタ21と加算器23との間に増幅器を設けることにより、入力音声信号とオールパスフィルタ21の出力との重みを変更することで、力み音声変換部分と非

変換部分とを切り替えるものとしても良い。

[0082] (実施の形態2の変形例)

図15は実施の形態2の力み変換部の変形例の機能ブロック図であり、図16は実施の形態2の力み変換部の変形例の動作を示すフローチャートである。図7および図8と同じ構成要素については同じ符号を用い、その詳細な説明は繰替えさない。

[0083] 図15に示されるように、本変形例の力み変換部20の構成は実施の形態2の図7に示す力み変換部20と同じであるが、実施の形態2において音声信号を入力として受け付けていたものを、音源波形を入力として受け付けるものとする。この変更に伴い、音声波形を生成するために音源波形によって駆動される声道フィルタ61が設けられている。

[0084] 次に、上記のような構成の力み変換部20の動作を図16に従って説明する。まず、力み変換部20は音源波形、音韻ラベル、ならびに音声の読み情報および韻律情報を取得する(ステップS61)。このとき、音韻ラベルは力み実時間範囲決定部12へ入力され、音声の読み情報および韻律情報は力み音韻位置決定部11へ入力される。また、音源波形は加算器23へ入力される。次に、力み音韻位置決定部11は、実施の形態2と同様に、読み情報と韻律情報とを力み易さ推定規則に当てはめて当該音韻の力み易さを求め、力み易さがあらかじめ定められた閾値を越えた場合に当該音韻を力み位置と決定する(ステップS2)。力み実時間範囲決定部12は、力み音韻位置決定部11で音韻単位で決定された力み位置と音韻ラベルとの対応をとり、音韻単位の力み音声の時間位置情報を音源波形上の時間範囲として特定し(ステップS63)、スイッチ22へ切り替え信号を出力する。一方、周期信号生成部13は、80Hzの正弦波を生成し(ステップS4)、オールパスフィルタ21へ出力する。オールパスフィルタ21は、周期信号生成部13より出力された80Hzの正弦波に従って位相シフト量を制御する(ステップS25)。スイッチ22は、入力される音源波形が力み実時間範囲決定部12より出力された「力み音声」で発声されるべき時間範囲に含まれる場合には(ステップS26でyes)、オールパスフィルタ21と加算器23とを接続し(ステップS27)、加算器23は、入力音源波形にオールパスフィルタ21の出力を加算し(ステップS78)、声道フィルタ61に出力する。一方、音源波形が力み実時間範囲決定部12より出力さ

れた「力み音声」で発声されるべき時間範囲に含まれない場合には(ステップS26でno)、スイッチ22はオールパスフィルタ21と加算器23との接続を遮断し、力み変換部20は入力音源波形をそのまま声道フィルタ61に出力する。声道フィルタ61は実施の形態1の変形例と同様に、力み変換部10に入力された音源波形に対応する声道フィルタを制御するための情報を入力として受け、振幅変調部14から出力される音源波形に対応する声道フィルタを形成する。振幅変調部14から出力された音源波形は声道フィルタ61を通過し、音声波形が生成される(ステップS67)。

[0085] かかる構成によれば、実施の形態2と同様、適切な位置に「力み」音声を発生させることで、発声器官の緊張の度合いを感じ取ることのできる、怒り、興奮、緊張、勢い込んだ話し方、または元気のある話し方のような、微細な時間構造を持った、質感のあるリアルな感情音声を生成することができる。また、オールパスフィルタによる位相変化を利用して振幅変調を行うことで、より複雑な振幅変化を生み出すことで聴感上の自然さを損ねず、聴取者は人工的な歪を感じにくい。さらに、実施の形態1の変形例と同様、主に口および舌の形状に関係する声道フィルタでは無く、音源波形に変調を掛けることによって、より実際の発声時の現象に近く、人工的な歪を感じにくいより自然な「力み」音声を生成することができる。

[0086] なお、本実施の形態ではステップS4において周期信号生成部13は80Hzの正弦波を出力し、オールパスフィルタ21の位相シフト量もこれに従うものとしたが、変動周波数は40Hz-120Hzの間のいずれの周波数でもよく、オールパスフィルタ21は正弦波以外の変動の特性を持つものとしても良い。

[0087] なお、本実施の形態ではスイッチ22はオールパスフィルタ21と加算器23との接続を切り替えるものとしたが、オールパスフィルタへの入力の入り、切りを切り替えるとしても良い。

[0088] なお、本実施の形態ではスイッチ22によってオールパスフィルタ21と加算器23との接続を切り替えることで力み音声変換部分と非変換部分とを切り替えたが、加算器23において、入力音声信号とオールパスフィルタ21の出力とに重みをつけて加算することにより力み音声変換部分と非変換部分とを切り替えるものとしてもよい。またはオールパスフィルタ21と加算器23との間に増幅器を設けることで入力音声信号とオ



ールパスフィルタ21の出力との重みを変更することで力み音声変換部分と非変換部分とを切り替えるものとしても良い。

[0089] (実施の形態3)

図17は実施の形態3の音声変換装置の構成を示す機能ブロック図である。図18は本実施の形態の同左を示すフローチャートである。図1および図10と同じ構成要素については同じ符号を用い、その詳細な説明は繰り返さない。

[0090] 図17に示されるように、本発明の音声変換装置は、入力された音声信号を力みのある音声信号に変換する装置であり、音素認識部31と、韻律分析部32と、力み範囲指定入力部33と、スイッチ34と、力み変換部10とを備えている。

[0091] 力み変換部10は、実施の形態1と同様であるので、その詳細な説明は繰り返さない。

[0092] 音素認識部31は、入力される音声を受付け、入力音声を音響モデルと照合し、音素列を出力する処理部である。

[0093] 韻律分析部32は、入力される音声を受付け、入力音声の基本周波数およびパワーを分析する処理部である。

[0094] 力み範囲指定入力部33は、使用者が力み音声に変換したい音声の範囲を指定する処理部である。例えば、力み範囲指定入力部33は、マイクロフォンまたは拡声器に設けられた「力みスイッチ」であり、使用者が力みスイッチを押し続ける間に入力された音声を「力み範囲」として指定する。または、力み範囲指定入力部33は、使用者が入力音声を監視しながら力み音声に変換したい音声が入力されている間「力みスイッチ」を押し続け、「力み範囲」を指定するための入力装置等である。

[0095] スイッチ34は、音素認識部31および韻律分析部32の出力を、力み音韻位置決定部11へ入力するか否かを切り替えるスイッチである。

[0096] 次に、上記の構成の音声変換装置の動作を図18に従って説明する。

[0097] まず、音声が入力される。このとき、入力音声は、音素認識部31と韻律分析部32とに入力される。音素認識部31は、入力された音声信号をスペクトル分析し、入力音声のスペクトル情報と音響モデルとを照合し、入力された音声の音素を決定する(ステップS31)。

- [0098] 一方、韻律分析部32は、入力された音声の基本周波数分析を行い、さらにパワーを求める(ステップS32)。スイッチ34は、力み範囲指定入力部33からの力み範囲指定入力があるか否かを判定する(ステップS33)。
- [0099] 力み範囲指定入力がある場合には(ステップS33でyes)、力み音韻位置決定部11は、読み情報と韻律情報とを力み易さ推定規則に当てはめて当該音韻の力み易さを求め、力み易さがあらかじめ定められた閾値を越えた場合に当該音韻を力み位置と決定する(ステップS2)。実施の形態1では数量化II類の独立変数のうち韻律情報として、アクセント核からの距離、またはアクセント句内での位置を用いる例を示したが、本実施の形態では韻律情報として基本周波数の絶対値、基本周波数の時間軸に対する傾き、またはパワーの時間軸に対する傾き等、韻律分析部32により分析された値を用いるものとする。
- [0100] 力み実時間範囲決定部12は、力み音韻位置決定部11で音韻単位で決定された力み位置と、音韻ラベルとの対応をとり、音韻単位の力み音声の時間位置情報を音声信号上の時間範囲として特定する(ステップS3)。
- [0101] 一方、周期信号生成部13は、80Hzの正弦波を生成し(ステップS4)、その正弦波信号に直流成分を加えた信号を生成する(ステップS5)。
- [0102] 振幅変調部14は、「力み位置」として特定された音声信号の実時間範囲について、周期信号生成部13が生成した80Hzで振動する周期信号を入力音声信号に乗ずることで、入力音声信号の振幅変調を行い(ステップS6)、音韻の時間長より周期が短い振幅の周期的変動を含む「力み」音声への変換を行い、力み音声を出力する(ステップS34)。
- [0103] 力み範囲指定入力がない場合は(ステップS33でno)、振幅変調部14は、入力音声信号を変形せずにそのまま出力する(ステップS29)。
- [0104] かかる構成によれば、入力音声のうち使用者の指定範囲において、音韻ごとの情報から推定規則に従ってその音韻を力み位置とするか否かを決定し、力み位置と推定された音韻についてのみ、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を行って、適切な位置に「力み」音声を発生させる。このことにより、入力音声を一様に変形する際にそれぞれ生じる、雑音が重畳しているような違和感および音質

が劣化したような印象を起こすことなく、入力音声に発声器官の緊張の度合いを感じ取ることのできる、怒り、興奮、緊張、勢い込んだ印象、または元気のある印象を、微細な時間構造を再現して、声の質感としてリアルに付け加え、音声をより一層表情豊かなものに変換することができる。すなわち、入力音が音声のみである場合にも、力み位置を推定するために必要な情報を抽出することができ、入力音声を適切な位置で「力み」音声を発声する表情豊かな音声に変換することができる。

[0105] なお、本実施の形態では、力み範囲指定入力部33によって制御され、かつ音素認識部31および韻律分析部32と力み音韻位置決定部11との接続を切り替えるスイッチ34によって、使用者の指定する範囲の音声についてのみ、力み音韻位置を決定するものとしたが、スイッチを音素認識部31と韻律分析部32との入力部分に置き換え、音素認識部31と韻律分析部32とへの音声信号の入力の入り、切りを切り替えるものとしても良い。

[0106] なお、本実施の形態では、力み変換部10によって力み音声への変換を行ったが、実施の形態2に示した力み変換部20によって力み音声への変換を行うものとしても良い。

[0107] (実施の形態3の変形例)

図19は実施の形態3の音声変換装置の変形例の機能ブロック図であり、図20は実施の形態3の音声変換装置の変形例の動作を示すフローチャートである。図9および図10と同じ構成要素については同じ符号を用い、その詳細な説明は繰り返さない。

[0108] 図19に示されるように、本変形例の音声変換装置の構成は実施の形態3の図9同様に力み範囲指定入力部33と、スイッチ34と、力み変換部10とを備える。本変形例の音声変換装置は、さらに、入力音声を受付け、ケプストラム分析を行う声道フィルタ分析部81と、声道フィルタ分析部より出力されるケプストラム係数に基づいて音素認識を行う音素認識部82と、声道フィルタ分析部より出力されるケプストラム係数に基づいて形成される逆フィルタ83と、逆フィルタ83によって抽出された音源波形より韻律を分析する韻律分析部84と、声道フィルタ61とを備える。

[0109] 次に、上記の構成の音声変換装置の動作を図20に従って説明する。まず、音声が入力される。このとき、入力音声は声道フィルタ分析部81に入力さ

れる。声道フィルタ分析部81は、入力された音声信号をケプストラム分析し、入力音声の声道フィルタを決定するケプストラム係数列を求める(ステップS81)。音素認識部82は、声道フィルタ分析部81より出力されたケプストラム係数と音響モデルとを照合し、入力された音声の音素を決定する(ステップS82)。一方、逆フィルタ83は、声道フィルタ分析部81より出力されたケプストラム係数を用いて逆フィルタを形成し、入力された音声の音源波形を生成する(ステップS83)。韻律分析部84は、逆フィルタ83より出力された音源波形の基本周波数分析を行い、さらにパワーを求める(ステップS84)。力み音韻位置決定部11は、力み範囲指定入力部33からの力み範囲指定入力があるか否かを判定する(ステップS33)。力み範囲指定入力がある場合には(ステップS33でyes)、力み音韻位置決定部11は、読み情報と韻律情報とを力み易さ推定規則に当てはめて、当該音韻の力み易さを求め、力み易さがあらかじめ定められた閾値を越えた場合に、当該音韻を力み位置と決定する(ステップS2)。力み実時間範囲決定部12は、力み音韻位置決定部11で音韻単位で決定された力み位置と音韻ラベルとの対応をとり、音韻単位の力み音声の時間位置情報を、音源波形上の時間範囲として特定する(ステップS63)。一方、周期信号生成部13は、80Hzの正弦波を生成し(ステップS4)、その正弦波信号に直流成分を加えた信号を生成する(ステップS5)。振幅変調部14は、「力み位置」として特定された音源波形の実時間範囲について、周期信号生成部13が生成した80Hzで振動する周期信号を音源波形に乗ずることで振幅変調を行う(ステップS66)。声道フィルタ61は、声道フィルタ分析部81より出力されたケプストラム係数列すなわち声道フィルタの制御情報に基づき声道フィルタを形成する。振幅変調部14から出力された音源波形は、声道フィルタ61を通過し、音声波形が生成される(ステップS67)。

- [0110] かかる構成によれば、入力音声のうち使用者の指定範囲において、音韻ごとの情報から推定規則に従ってその音韻を力み位置とするか否かを決定し、力み位置と推定された音韻についてのみ、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を行って、適切な位置に「力み」音声を発生させることで、入力音声を一様に変形する際に生じる雑音が重畳しているような違和感や音質が劣化したような印象を起こすことなく、入力音声に発声器官の緊張の度合いを感じ取ることのできる、怒り、興

奮、緊張、勢い込んだ印象、または元気のある印象を、微細な時間構造を再現して、声の質感としてリアルに付け加え、音声をより一層表情豊かなものに変換することができる。すなわち、入力が音声のみである場合にも力み位置を推定するために必要な情報を抽出することができ、入力音声を適切な位置で「力み」音声を発声する表情豊かな音声に変換することができる。さらに、実施の形態1の変形例と同様、主に口や舌の形状に関係する声道フィルタでは無く、音源波形に変調を掛けることでより実際の発声時の現象に近く、人工的な歪を感じにくいより自然な「力み」音声を生成することができる。

[0111] なお、本実施の形態では、力み範囲指定入力部33によって制御され、音素認識部82および韻律分析部84と力み音韻位置決定部11との接続を切り替えるスイッチ34によって使用者の指定する範囲の音声についてのみ力み音韻位置を決定するものとしたが、スイッチを音素認識部82と韻律分析部84との入力部分に置き換え、音素認識部82と韻律分析部84への入力の入り、切りを切り替えるものとしても良い。

[0112] なお、本実施の形態では、力み変換部10によって力み音声への変換を行ったが、実施の形態2およびその変形例に示した力み変換部20によって力み音声への変換を行うものとしても良い。

[0113] (実施の形態4)

図21は、実施の形態4の音声合成装置の構成を示す機能ブロック図である。図22は本実施の形態の同左を示すフローチャートである。図23は本実施の形態の1変形例の音声合成装置の構成を示す機能ブロック図である。図24および図25は変形例の音声合成装置への入力例を示す図である。図21および図22について、図1および図10と同じ構成要素については同じ符号を用い、その詳細な説明は繰り返さない。

[0114] 図21に示されるように、本発明の音声合成装置は、入力されたテキストを読み上げた音声を合成する装置であり、テキスト入力部40と、言語処理部41と、韻律生成部42と、波形生成部43と、力み範囲指定入力部44と、力み音韻位置指定部46と、切り替え入力部47と、スイッチ45と、スイッチ48と、力み変換部10とを備えている。

[0115] 力み変換部10は、実施の形態1と同様であるので、その詳細な説明は繰り返さない。

い。

- [0116] テキスト入力部40は、使用者の入力によるテキストまたは他の方法によるテキストの入力を受け、言語処理部41およびカミ範囲指定入力部44に出力する処理部である。
- [0117] 言語処理部41は、入力テキストを受け、形態素解析によってテキストを単語に分割してその読みを特定し、さらに構文解析により単語間の係り受け関係を明らかにして単語の読みの変形を行いアクセント句やフレーズといった記述的韻律情報を生成する処理部である。
- [0118] 韻律生成部42は、言語処理部41より出力された読みと記述的韻律情報とより、各音韻およびポーズの時間長、基本周波数、振幅またはパワーの値を生成する処理部である。
- [0119] 波形生成部43は、言語処理部41より出力された読み情報と韻律生成部42より出力された音韻およびポーズの時間長、基本周波数、振幅またはパワーの値を受け、指定された音声波形を生成する処理部である。波形生成部43は、波形接続型の音声合成方式であれば、素片選択部と素片データベースとを備える。また、波形生成部43は、規則合成型の音声合成方式であれば、採用する生成モデルに合わせて、生成モデルと信号生成部とを備える。
- [0120] カミ範囲指定入力部44は、使用者がカミ音声で発声させたいテキスト上の範囲を指定する処理部である。例えば、使用者が入力したテキストをディスプレイ(display)に表示し、テキストの表示をポインティングすることで反転させてテキスト上で「カミ範囲」を指定するための入力装置等である。
- [0121] カミ音韻位置指定部46は、使用者がカミ音声で発声させたい範囲を音韻単位で指定する処理部である。例えば言語処理部41が出力した音韻列をディスプレイに表示し、表示された音韻列をポインティングすることで反転させて音韻単位で「カミ位置」を指定するための入力装置等である。
- [0122] 切り替え入力部47は、カミ音韻位置を使用者が設定する方法と自動で設定する方法とを切り替える入力を受け、スイッチ48を制御する処理部である。
- [0123] スイッチ45は、スイッチ48を経由して言語処理部41とカミ音韻位置決定部11との

接続を切り替えるスイッチであり、スイッチ48は、力み音韻位置決定部11への入力を言語処理部41の出力と、力み音韻位置指定部46からの使用者の入力との間で切り替えるスイッチである。

- [0124] 次に、上記の構成の音声合成装置の動作を図22に従って説明する。
- [0125] まず、テキスト入力部40は、入力テキストを受付ける(ステップS41)。テキストの入力とは、例えばキーボードからの入力、既に記録されたテキストデータの入力、および文字認識による読み込み等である。テキスト入力部40は、入力テキストを言語処理部41および力み範囲指定入力部44に出力する。
- [0126] 言語処理部41は、形態素解析および構文解析により音韻列と記述的韻律情報とを生成する(ステップS42)。形態素解析および構文解析では、例えばNgramのような、言語モデルと辞書とを用い、入力テキストとモデルとのマッチングを取ることで、最適な単語分割を行い、各単語の係り受けの解析を行う。また、単語の読みと単語間の係り受けに基づき、アクセント、アクセント句、フレーズといった記述的韻律情報を生成する。
- [0127] 韻律生成部42は、言語処理部41より出力された音韻情報と記述的韻律情報とを取得し、音韻列と記述的韻律情報とに基づいて各音韻およびポーズの時間長、基本周波数、パワーまたは振幅の値を決定する(ステップS43)。韻律の数値情報の生成は、例えば、統計的学習によって作られた韻律生成モデル、または発声機構から導かれた韻律生成モデルに基づいて行なわれる。
- [0128] 波形生成部43は、言語処理部41から出力された音韻情報と韻律生成部42より出力された韻律数値情報とを受付け、対応する音声波形を生成する(ステップS44)。波形生成方法としては、例えば、音韻列と韻律情報とに従って最適な音声素片を選択して接続する波形接続による方法、韻律情報に従って音源信号を生成し、音韻列に従って設定した声道フィルタを通過させることで音声波形を生成する方法、および、音韻列と韻律情報とから、スペクトルパラメータを推定して音声波形を生成する方法等がある。
- [0129] 一方、力み範囲指定入力部44は、ステップS41で入力されたテキストを取得し、使用者に提示する(ステップS45)。また、力み範囲指定入力部44は、使用者がテキス

ト上で指定する力み範囲を取得する(ステップS46)。

- [0130] 力み範囲指定入力部44に入力テキストの全部または一部を指定する入力がない場合(ステップS47でNo)、力み範囲指定入力部44はスイッチ45を切断し、本実施の形態の音声合成装置はステップS44で生成された合成音声を入力する(ステップS53)。
- [0131] 力み範囲指定入力部44に入力テキストの全部または一部を指定する入力がある場合には(ステップS47でYes)、力み範囲指定入力部44は、入力テキスト中の力み範囲を特定し、スイッチ45を接続することにより、言語処理部41より出力された音韻情報・記述的韻律情報および力み範囲情報をスイッチ48へ接続する。また、言語処理部41より出力された音韻列は、力み音韻位置指定部46へ出力され、使用者に提示される(ステップS49)。
- [0132] 力み範囲としての大まかな範囲指定ではなく、力み音韻位置を詳細に指定したい使用者は、力み音韻位置を手動で入力指定することができるようにするために、切り替え入力部47に切り替え入力をする。
- [0133] 力み音韻位置指定への切り替え入力がある場合(ステップS50でYes)、切り替え入力部47は、スイッチ48を力み音韻位置指定部46に接続する。力み音韻位置指定部46は、使用者の力み音韻位置指定情報を受付ける(ステップS51)。使用者は、例えばディスプレイに提示された音韻列上で力み音声で発声されるべき音韻を指定することにより、力み音韻位置を指定する。
- [0134] 力み音韻位置指定入力がない場合(ステップS52でNo)、力み音韻位置決定部11は力み音韻位置としていずれの音韻も指定せず、本実施の形態の音声合成装置はステップS44で生成された合成音声を入力する(ステップS53)。
- [0135] 一方、力み音韻位置指定入力がある場合(ステップS52でYes)、力み音韻位置決定部11は、ステップS51で力み音韻位置指定部46より入力された音韻位置を力み音韻位置として決定する。
- [0136] 力み音韻位置指定への切り替え入力がない場合は(ステップS50でNo)、力み音韻位置決定部11は、ステップS48で特定された力み範囲に対して実施の形態1と同様に、音韻ごとに、音声の読み情報および韻律情報を「力み易さ」推定式に当てはめ



、各音韻の「力み易さ」を求める。また、力み音韻位置決定部11は、求めた「力み易さ」があらかじめ定められた閾値を越えた音韻を「力み位置」として決定する(ステップS2)。実施の形態1では数量化II類を用いる例を示したが、本実施の形態では音韻情報および韻律情報を入力とするSVM(Support Vector Machine)を用いて、音声がかんでいたか、かんでいなかったかの2クラスのカテゴリを予測するものとする。SVMについても他の統計的手法と同様、「力み」音声を含む学習用音声データについて、音韻ごとに当該音韻、直前音韻、直後音韻、アクセント句内位置、ならびにアクセント核に対する相対位置、フレーズ内位置および文中位置を入力とし、当該音声が力み音声であったか否かを推定するモデルが学習される。力み音韻位置決定部11は、言語処理部41が出力した音韻情報および記述的韻律情報より、SVMの入力変数である当該音韻、直前音韻、直後音韻、アクセント句内位置、ならびにアクセント核に対する相対位置、フレーズ内位置および文中位置を抽出し、各音韻が力み音声中で発生されるべきか否かを決定する。

- [0137] 力み実時間範囲決定部12は、「力み位置」として決定された音韻の時間位置情報を、韻律生成部42が出力する各音韻の時間長情報すなわち音韻ラベルに基づいて、波形生成部43が出力する合成音声波形上の時間範囲として特定する(ステップS3)。
- [0138] 実施の形態1と同様に、周期信号生成部13は、周波数80Hzの正弦波を生成し(ステップS4)、正弦波に直流成分を加算する(ステップS5)。
- [0139] 振幅変調部14は、「力み位置」として特定された音声信号の時間範囲について、合成音声信号に直流成分を加算した周期成分を乗じる(ステップS6)。本実施の形態の音声合成装置は、力み音声を含んだ合成音声出力する(ステップS34)。
- [0140] かかる構成によれば、入力テキストのうち使用者の指定範囲において、音韻ごとの情報から推定規則に従ってその音韻を力み位置とするか否かを決定し、力み位置と推定された音韻についてのみ、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を行って、適切な位置に「力み」音声を発生させる。または入力テキストを音声に変換する際の音韻列のうち使用者の指定する音韻において、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を行って、「力み」音声を発生させる。この

ことにより、入力音声を一様に変形する際に生じる雑音が重畳しているような違和感および音質が劣化したような印象を起こすことがなくなる。また、使用者の自由なデザインにより、発声器官の緊張の度合いを感じ取ることのできる、怒り、興奮、緊張、勢い込んだ印象または元気のある印象を、微細な時間構造として再現し、かつ、これらを声の質感として、入力音声にリアルに付け加えて詳細に音声の表情をつくりこむことができる。すなわち、変換元となる音声入力がない場合にも、入力テキストより合成音声を生成し、変換元となる音声を生成することで、適切な位置で「力み」音声を発声する表情豊かな音声に変換することができる。さらには、「力み」音声による素片データベースおよび合成パラメータデータベースを必要とせず、簡単な信号処理のみで力み音声を生成することができる。このため、データ量および計算量を大幅に増大させることなく、発声器官の緊張の度合いを感じ取ることのできる、怒り、興奮、緊張、勢い込んだ話し方または元気のある話し方のような、微細な時間構造による質感のあるリアルな感情音声を生成することができる。

[0141] なお、本実施の形態では、力み範囲指定入力部44を用いて使用者がテキスト上で力み範囲を指定することにより力み範囲を入力し、入力されたテキスト上の範囲に対応する合成音声の中で力み音韻位置を決定し、力み音声を発生させるものとしたが、この方法に限定されるものではない。例えば、図24のような力み範囲を示すタグ情報が付随したテキストを入力として受け、力み範囲指定取得部51がタグ情報と合成音声に変換すべきテキストの情報とを分離し、タグ情報を解析してテキスト上での力み範囲指定情報を取得するものとしても良い。また、「力み音韻位置指定部46」の入力についても、例えば図24および図25に示す、特許文献：特開2006-227589号公報に記載のような形式により、音韻ごとに力み音声で発声するか否かを指定するタグによって指定するものとしても良い。図24のタグ情報は<voice>タグに囲まれた領域のテキストを合成する際の音声について、「quality(声質)」を「力み」で合成することを指定するものである。すなわち「あらゆる現実をすべて自分の方へ捻じ曲げたのだ。」というテキストのうち「捻じ曲げたのだ」の範囲について「力み」音声を指定するものである。図25のタグ情報は、<voice>タグで囲まれた範囲のうち、先頭から5モーラの音韻について「力み」音声を指定するものである。

- [0142] なお、本実施の形態では力み音韻位置決定部11は言語処理部41より出力された音韻情報とアクセント等の記述的韻律情報とを用いて力み音韻位置を推定するものとしたが、韻律生成部42を言語処理部41と同様にスイッチ45に接続し、スイッチ45は言語処理部41と韻律生成部42の出力を力み音韻位置決定部11に接続するものとしても良い。これにより、力み音韻位置決定部11は、言語処理部41より出力された音韻情報と、韻律生成部42より出力された基本周波数やパワーの数值情報を用いて、実施の形態3のように音韻情報と物理量としての韻律情報である基本周波数やパワーの値とを用いて力み音韻位置を推定するものとしても良い。
- [0143] なお、本実施の形態では力み音韻位置を使用者が指定するために切り替え入力部47を設けてスイッチ48を切り替えたが、力み音韻位置指定部46に入力がある場合にスイッチを切り替えるものとしても良い。
- [0144] なお、本実施の形態ではスイッチ48は力み音韻位置決定部11への入力を切り替えるものとしたが、力み音韻位置決定部11から力み実時間範囲決定部12への接続を切り替えるものとしても良い。
- [0145] なお、本実施の形態では、力み変換部10によって力み音声への変換を行ったが、実施の形態2に示した力み変換部20によって力み音声への変換を行うものとしても良い。
- [0146] なお、実施の形態3の力み範囲指定入力部33および実施の形態4の力み範囲指定入力部44は、力み音声を発声させる範囲を指定するものとしたが、音声を力ませない範囲を指定するものとしても良い。
- [0147] なお、本実施の形態では、韻律生成部42は言語処理部41より出力された読みと記述的韻律情報より、各音韻およびポーズの時間長、基本周波数、振幅またはパワーの値を生成するものとしたが、読みと記述的韻律情報に加えて、力み範囲指定入力部44の出力を受け、力み範囲について基本周波数のダイナミックレンジを大きくし、さらにパワーまたは振幅の平均値を大きくし且つダイナミックレンジを大きくするものとしても良い。これにより、変換元の音声を「力み」音声が発生する音声としてよりふさわしい、力を入れて発声しているものにすることができ、より一層質感のあるリアルな感情表現が可能となる。

[0148] (実施の形態4の他の変形例)

図26は実施の形態4の音声合成装置の他の変形例の機能ブロック図であり、図27は実施の形態4の音声合成装置の他の変形例の動作を示すフローチャートである。図13および図14と同じ構成要素については同じ符号を用い、その詳細な説明は繰り返さない。

[0149] 図26に示されるように、本変形例の音声変換装置の構成は、実施の形態4の図13と同様に、テキスト入力部40と、言語処理部41と、韻律生成部42と、力み範囲指定入力部44と、力み音韻位置指定部46と、切り替え入力部47と、スイッチ45と、スイッチ48と、力み変換部10とを備える。また、本変形例の音声変換装置は、波形接続によって音声波形を生成する波形生成部43に替わり、音源波形を生成する音源波形生成部93、声道フィルタの制御情報を生成するフィルタ制御部94および声道フィルタ61を備える。

[0150] 次に、上記の構成の音声変換装置の動作を図27に従って説明する。まず、テキスト入力部40は、入力テキストを受付け(ステップS41)、入力テキストを言語処理部41および力み範囲指定入力部44に出力する。言語処理部41は、形態素解析および構文解析により音韻列と記述的韻律情報を生成する(ステップS42)。韻律生成部42は、言語処理部41より出力された音韻情報と記述的韻律情報とを取得し、音韻列と記述的韻律情報とに基づいて各音韻およびポーズの時間長、基本周波数、パワーまたは振幅の値を決定する(ステップS43)。音源波形生成部93は、言語処理部41より出力された音韻情報と韻律生成部42より出力された韻律数値情報とを受付け、それらに対応する音源波形を生成する(ステップS94)。音源波形は、例えば、Rosenberg-Klattモデル(非特許文献:Klatt, D. and Klatt, L. "Analysis, synthesis, and perception of voice quality variations among female and male talkers", J. Acoust. Soc. Amer. Vol. 87, 820-857, 1990)のような音源モデルの制御パラメータを、音韻および韻律数値情報に対応して生成することにより生成される。音源モデルのパラメータのうち声門開放度および音源スペクトル傾斜等を用いた音源波形の生成方法としては、基本周波数、パワー、振幅、音声の継続時間長および音韻から統計的に上記パラメータを推定することにより音源波形を生成する方法、または、

自然音声から抽出された音源波形を記録したデータベースより音韻と韻律情報とに従って最適な音源波形を選択して接続する方法等がある。フィルタ制御部94は、言語処理部41より出力された音韻情報と韻律生成部42より出力された韻律数値情報とを受付け、それらの情報に対応するフィルタ制御情報を生成する(ステップS95)。声道フィルタの決定方法としては、例えば、音韻に従って複数のバンドパスフィルタの中心周波数と帯域とを設定する方法、または、ケプストラム係数若しくはスペクトルを音韻、基本周波数およびパワー等より統計的に推定し、それに従ってフィルタの係数を設定する方法等がある。一方、力み範囲指定入力部44は、ステップS41で入力されたテキストを取得し、使用者に提示する(ステップS45)。力み範囲指定入力部44は、使用者がテキスト上で指定する力み範囲を取得する(ステップS46)。力み範囲指定入力部44に入力テキストの全部または一部を指定する入力がない場合(ステップS47)、力み範囲指定入力部44は、スイッチ45を切断し、声道フィルタ61は、ステップS95で設定されたフィルタ制御情報に基づき声道フィルタを形成する。声道フィルタ61は、ステップS94で生成された音源波形から音声波形を生成する(ステップS67)。ステップS47において力み範囲指定入力部44に入力テキストの全部または一部を指定する入力がある場合は(ステップS47でYes)、力み範囲指定入力部44は、入力テキスト中の力み範囲を特定し、スイッチ45を接続して言語処理部41より出力された音韻情報・記述的韻律情報および力み範囲情報をスイッチ48へ出力する(ステップS48)。また、言語処理部41より出力された音韻列は力み音韻位置指定部46へ出力され、使用者に提示される(ステップS49)。力み音韻位置を詳細に指定したい使用者は、力み音韻位置を手動で入力指定することができるようにするために、切り替え入力部47に切り替えの入力をする。

[0151] 力み音韻位置指定への切り替え入力がある場合(ステップS50)、切り替え入力部47は、スイッチ48を力み音韻位置指定部46に接続し、使用者の力み音韻位置指定情報を受付ける(ステップS51)。力み音韻位置指定入力がない場合(ステップS52でNo)、力み音韻位置決定部11は、力み位置としていずれの音韻も指定せず、声道フィルタ61は、ステップS95で設定されたフィルタ制御情報に基づき声道フィルタを形成する。声道フィルタ61は、ステップS94で生成された音源波形から音声波形を

生成する(ステップS67)。一方、ステップS52において力み音韻位置指定入力がある場合(ステップS52でYes)、力み音韻位置決定部11は、ステップS51で力み音韻位置指定部46より入力された音韻位置を力み音韻位置として決定する(ステップS63)。ステップS50で力み音韻位置指定への切り替え入力がない場合は(ステップS50でNo)、力み音韻位置決定部11は、ステップS48で特定された力み範囲に対して音韻ごとに、音声の読み情報および韻律情報を「力み易さ」推定式に当てはめ、各音韻の「力み易さ」を求め、「力み易さ」があらかじめ定められた閾値を越えた音韻を「力み位置」として決定する(ステップS2)。力み実時間範囲決定部12は、「力み位置」として決定された音韻の時間位置情報を、韻律生成部42が出力する各音韻の時間長情報すなわち音韻ラベルに基づいて、音源波形生成部93が出力する合成音声波形上の時間範囲として特定する(ステップS63)。周期信号生成部13は、周波数80Hzの正弦波を生成し(ステップS4)、正弦波に直流成分を加算する(ステップS5)。振幅変調部14は、「力み位置」として特定された音源波形の時間範囲について、音源波形に周期成分を乗ずる(ステップS66)。声道フィルタ61はステップS95で設定されたフィルタ制御情報に基づき声道フィルタを形成し、ステップS66で「力み位置」の振幅を変調された音源波形を通過させて音声波形を生成する(ステップS67)。

[0152] かかる構成によれば、入力テキストのうち使用者の指定範囲において、音韻ごとの情報から推定規則に従ってその音韻を力み位置とするか否かを決定し、力み位置と推定された音韻についてのみ、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を行って、適切な位置に「力み」音声を発生させる、または入力テキストを音声に変換する際の音韻列のうち使用者の指定する音韻において、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を行って、「力み」音声を発生させる。このことにより、入力音声を一様に変形する際に生じる雑音が重畳しているような違和感および音質が劣化したような印象を起こすことがなくなる。また、使用者の自由なデザインにより、発声器官の緊張の度合いを感じ取ることのできる、怒り、興奮、緊張、勢い込んだ印象または元気のある印象を、微細な時間構造として再現し、かつ、これらを声の質感として、入力音声にリアルに付け加えて詳細に音声の表情をつくりこむことができる。すなわち、変換元となる音声入力がない場合にも、入力テキストより合成

音声を生成し、変換元となる音声を生成することで、適切な位置で「力み」音声を発声する表情豊かな音声に変換することができる。さらには、「力み」音声による素片データベースおよび合成パラメータデータベースを必要とせず、簡単な信号処理のみで力み音声を生成することができる。このため、データ量および計算量を大幅に増大させることなく、発声器官の緊張の度合いを感じ取ることのできる、怒り、興奮、緊張、勢い込んだ話し方または元気のある話し方のような、微細な時間構造を持った、質感のあるリアルな感情音声を生成することができる。加えて、本変形例によれば、実施の形態3の変形例と同様、主に口や舌の形状に関する声道フィルタではなく、音源波形に変調を掛けることで、より実際の発声時の現象に近く、人工的な歪を感じにくい、より自然な「力み」音声を生成することができる。

[0153] なお、実施の形態1、2および3では力み音韻位置決定部11は数量化II類による推定規則を用いるとし、実施の形態4ではSVMによる推定規則を用いるとしたが、実施の形態1、2、および3においてSVMによる推定規則を用い、実施の形態4で数量化II類による推定規則を用いても良い。さらに、ニューラルネット等これ以外の方法に基づいた推定規則を用いるものとしても良い。

[0154] なお、実施の形態3ではリアルタイムに音声に力みを付与するものとしたが、録音音声を用いるものとしても良い。さらには、実施の形態4のように力み音韻位置指定部を備え、あらかじめ音素認識を行った録音音声に対して、力み音声に変換する音韻を使用者が指定するものとしても良い。

[0155] なお、実施の形態1、3および4では、周期信号生成部13は80Hzの周期信号を生成するものとしたが、「力み」として聴取できる40Hzから120Hzの間でランダムな周期変動を持つ周期信号を生成してもよい。歌唱において、旋律にあわせて母音の時間長を長く伸ばす場合が多くあるが、時間長の長い(例えば、3秒を越える)母音に一定の変動周波数で振幅変動を付加すると、音声と共にブザー音が聞こえる等の不自然な音が生成される場合がある。振幅変動の変動周波数をランダムに変化させることでブザー音または雑音の重畳のような印象を減らせる場合もある。そこで、変動周波数をランダムに変化させることで、より実際の音声の振幅変動に近づき、自然な音声を生成することができる。

[0156] 今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は上記した説明ではなくて請求の範囲によって示され、請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される。

#### 産業上の利用可能性

[0157] 本発明にかかる音声変換装置および音声合成装置は、人間が怒鳴ったとき、強調のために力を入れて話すとき、興奮または緊張して話すとき等に現れる、だみ声、荒れた声、ざらざらした声 (harsh voice) や、演歌等の歌唱時に現れる「こぶし」や「うなり」のような表現、ブルースやロック等の歌唱時に現れる「シャウト」のような表現を行なう際に発生する、通常の発声とは異なる特徴を持った音声である「力み」音声を、力み音声用素片データベースおよび力み音声用パラメータデータベースを持つことなく、音韻の時間長より周期が短い周期的な振幅変動を含む変調を行うという単純な構成で生成することができる。また、その「力み」音声を音声中の適切な位置に生成することができる。このため、話者の発声器官の緊張や力の入り具合を、微細な時間構造を再現して、声の質感としてリアルに感じさせる表情豊かな音声を生成することができる。また、「力み」音声を音声中のどこに発生させるかを使用者がデザインすることもでき、音声の表情を詳細に調節して作りこむことができる。このような特徴を持つことで、カーナビゲーションシステム、テレビジョン受像機、オーディオシステム等の電子機器、またはロボット等の音声・対話インタフェース等として有用である。

[0158] 本発明をカラオケに用いることもできる。例えば、マイクロフォンに「力み」変換スイッチを設け、歌い手が当該スイッチを押すことにより、入力音声に「力み」、「うなり」または「こぶし」といった表現をつけ加えることができる。さらにカラオケのマイクの握りに圧力センサやジャイロセンサをつけることで、歌い手が力を込めて歌っていることを検出し、検出結果に応答して声に自動的に表現をつけることも可能である。このようにして歌声に表現を付け加え、歌う楽しみを増大させることができる。

[0159] また、本発明を拡声器に用いれば演説や講演時に、強調したいところは「力み」音声への変換を指定し、力が入った説得力のある話し方にすることができる。

[0160] また、本発明を電話に利用すれば、迷惑電話に対して自分の声を「力み」音声に変



換して相手へ送信することで、所謂「どすの効いた声」により迷惑電話を撃退することにも利用可能である。同様に、本発明をインターホンに用いれば、迷惑な来訪者を撃退することにも利用可能である。

- [0161] 本発明をラジオに利用すれば、強調したい単語またはジャンル等を登録しておき、使用者が興味のある情報を「力み」音声に変換して出力することで強調し、使用者が情報を聞き逃さないようにすることもできる。また、コンテンツ流通においては、同一コンテンツであっても利用者の特性や状況に応じて、「力み」範囲を変え、利用者にあった情報の訴求ポイントを強調することにも利用できる。
- [0162] 本発明を施設内の音声案内に用いれば、案内内容の危険度、緊急度または重要度に合わせて「力み」を付加し、聴取者の注意を引くことにも利用可能である。
- [0163] さらに本発明を機器の内部状態を示す音声出力インタフェースに利用すれば、機器の稼働状態が高い場合や、計算量が大きい場合等の状況で音声を出力する際に「力み」を付加し、機器が「頑張っている」ように表現することでより親しみの持てるインタフェースを設計することにも利用可能である。

## 請求の範囲

- [1] 変換の対象となる音声の中の音韻を指定する力み音韻位置指定手段と、  
前記力み音韻位置指定手段により指定された音韻を表す音声波形に、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を施す変調手段とを備える  
ことを特徴とする力み変換装置。
- [2] 前記変調手段は、前記力み音韻位置指定手段により指定された音韻を表す音声波形に、40Hz以上の周波数の周期的な振幅変動を伴う変調を施す  
ことを特徴とする請求項1に記載の力み変換装置。
- [3] 前記変調手段は、前記力み音韻位置指定手段により指定された音韻を表す音声波形に、40Hz以上でかつ120Hz以下の周波数の周期的な振幅変動を伴う変調を施す  
ことを特徴とする請求項1に記載の力み変換装置。
- [4] 前記変調手段は、前記力み音韻位置指定手段により指定された音韻を表す音声波形に、振幅の変動幅を百分率で定義した周期的な振幅変動の変調度が40%以上でかつ80%以下となる、周期的な振幅変動を伴う変調を施す  
ことを特徴とする請求項1に記載の力み変換装置。
- [5] 前記変調手段は、周期信号を音声波形に乗ずることにより周期的な振幅変動を伴う変調を前記音声波形に施す  
ことを特徴とする請求項1に記載の力み変換装置。
- [6] 前記変調手段は、  
前記力み音韻位置指定手段により指定された音韻を表す音声波形の位相をずらすオールパスフィルタと、  
前記力み音韻位置指定手段により指定された音韻を表す音声波形に、前記オールパスフィルタにより位相がずらされた音声波形を加算する加算手段とを備える  
ことを特徴とする請求項1に記載の力み変換装置。
- [7] さらに、前記力み音韻位置指定手段によって指定される、変換の対象となる音声の中の音韻を含み得る音声の範囲を指定する力み範囲指定手段を備える  
ことを特徴とする請求項1に記載の力み変換装置。

- [8] 音声波形を受付ける入力手段と、  
力み音声に変換すべき音韻を指定する力み音韻位置指定手段と、  
前記力み音韻位置指定手段による力み音声に変換すべき音韻の指定に従って、  
前記入力手段で受け付けられた音声波形に、音韻の時間長より周期が短い周期的な  
振幅変動を伴う変調を施す変調手段とを備える  
ことを特徴とする音声変換装置。
- [9] さらに、前記力み音韻位置指定手段によって指定される、変換の対象となる音韻を  
含む得る音声の範囲を指定する力み範囲指定入力手段を備える  
ことを特徴とする請求項8に記載の音声変換装置。
- [10] さらに、  
前記音声波形の音韻列を認識する音韻認識手段と、  
前記音声波形の韻律情報を抽出する韻律分析手段とを備え、  
前記力み音韻位置指定手段は、前記音韻認識手段により認識された入力音声の  
音韻列と前記韻律分析手段により抽出された韻律情報とに基づいて力み音声に変  
換すべき音韻を指定する  
ことを特徴とする請求項8に記載の音声変換装置。
- [11] 使用者による力み音声に変換すべき音韻を指定する入力を受け付ける力み音韻  
位置入力手段を備え、  
前記変調手段は、前記力み音韻位置入力手段による音声に変換すべき音韻の指  
定に従って、前記入力手段で受け付けられた音声波形に周期的な振幅変動を伴う変  
調を施す  
ことを特徴とする請求項8に記載の音声変換装置。
- [12] テキストを受付ける入力手段と、  
前記入力手段で受け付けられた前記テキストを解析して読み情報と韻律情報とを生  
成する言語処理手段と、  
前記読み情報と韻律情報とに従って音声波形を生成する音声合成手段と、  
力み音声に変換すべき音韻を指定する力み音韻位置指定手段と、  
前記音声合成手段で合成された音声波形のうち、前記力み音韻位置指定手段より

指定された音韻を表す音声波形に、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を施す変調手段とを備える

ことを特徴とする音声合成装置。

- [13] さらに、前記力み音韻位置指定手段によって指定される力み音声を生成すべき音韻を含み得る範囲を指定する力み範囲指定入力手段を備える

ことを特徴とする請求項12に記載の音声合成装置。

- [14] 前記入力手段は、変換すべき内容と共に、合成する音声の特性を指定する情報を含み、かつ当該情報は前記力み音声を生成すべき音韻を含みうる範囲の情報を含む、テキストを受け、

前記音声合成装置は、前記入力手段で受け付けられた前記テキストを解析して前記力み音声を生成すべき音韻を含みうる範囲を取得する力み範囲指定取得手段を備える

ことを特徴とする請求項12に記載の音声合成装置。

- [15] 前記力み音韻位置指定手段は、前記言語処理手段により生成された読み情報と韻律情報とに基づいて力み音声に変換すべき音韻を指定する

ことを特徴とする請求項12に記載の音声合成装置。

- [16] 前記力み音韻位置指定手段は、前記言語処理手段により生成された読み情報と前記音声合成手段により生成された音声波形の基本周波数、パワー、振幅、音韻時間長の少なくとも何れか一つとに基づいて、力み音声に変換すべき音韻を指定する

ことを特徴とする請求項12に記載の音声合成装置。

- [17] さらに、使用者による力み音声に変換すべき音韻を指定する入力を受け付ける力み音韻位置入力手段を備え、

前記変調手段は、前記音声合成手段で合成された音声波形のうち、前記力み音韻位置入力手段により指定された音韻を表す音声波形に、周期的な振幅変動を伴う変調を施す

ことを特徴とする請求項12に記載の音声合成装置。

- [18] 変換の対象となる音声中の部分を音韻単位で指定し、

前記指定された音韻位置において、音韻の時間長より周期が短い周期的な振幅変

動を伴う変調を音声波形に施す

ことを特徴とする音声変換方法。

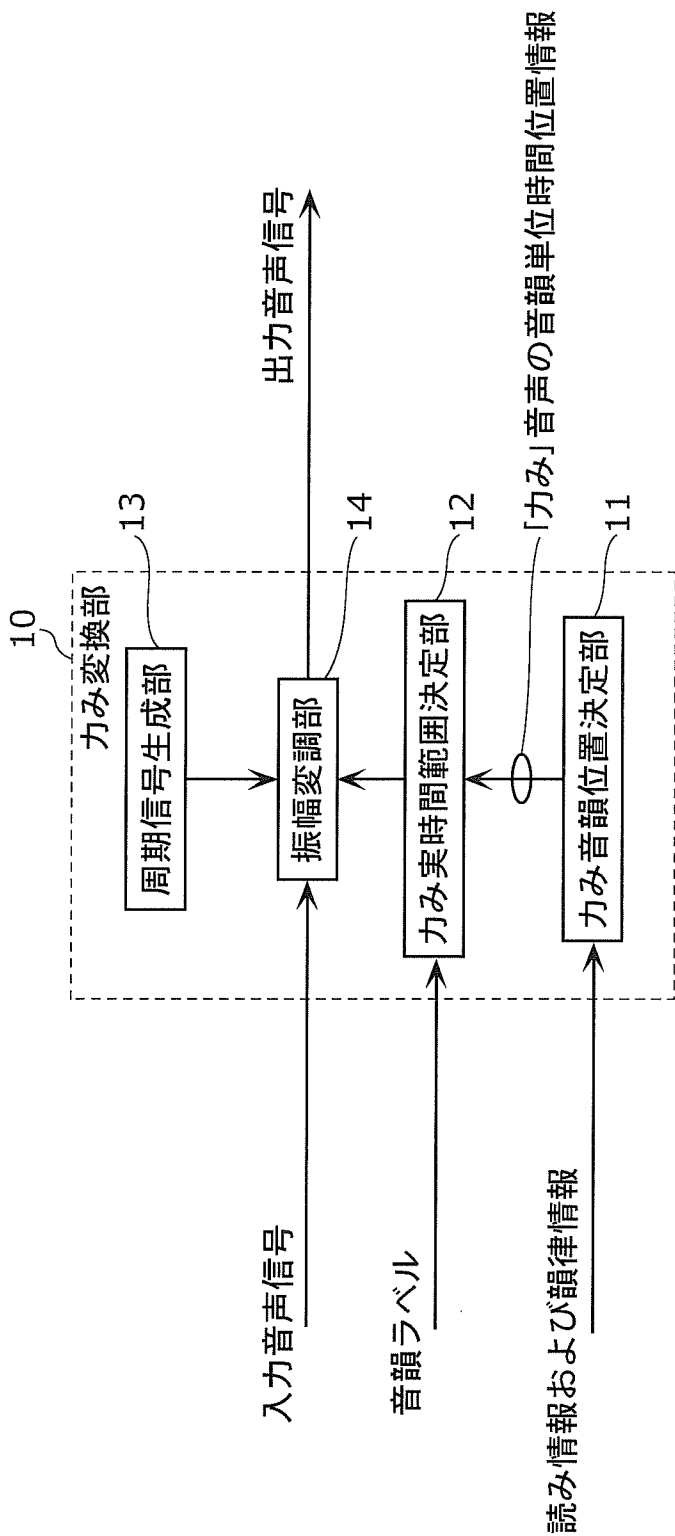
- [19] 力み音声を生成すべき音韻を指定し、  
前記指定された音韻位置において、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を音声波形に施した合成音声を生成する  
ことを特徴とする音声合成方法。
- [20] 変換の対象となる音声中の部分を音韻単位で指定するステップと、  
前記指定された音韻位置において、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を音声波形に施すステップとをコンピュータに実行させる  
ことを特徴とする音声変換プログラム。
- [21] 力み音声を生成すべき音韻を指定するステップと、  
前記指定された音韻位置において、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を音声波形に施した合成音声を生成するステップとをコンピュータに実行させる  
ことを特徴とする音声合成プログラム。
- [22] 音声変換プログラムを記録したコンピュータ読取可能な記録媒体であって、  
前記音声変換プログラムは、  
変換の対象となる音声中の部分を音韻単位で指定するステップと、  
前記指定された音韻位置において、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を音声波形に施すステップとをコンピュータに実行ためのプログラムである  
ことを特徴とするコンピュータ読取可能な記録媒体。
- [23] 音声合成プログラムを記録したコンピュータ読取可能な記録媒体であって、  
前記音声合成プログラムは、  
力み音声を生成すべき音韻を指定するステップと、  
前記指定された音韻位置において、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を音声波形に施した合成音声を生成するステップとをコンピュータに実行させるためのプログラムである

ことを特徴とするコンピュータ読取可能な記録媒体。

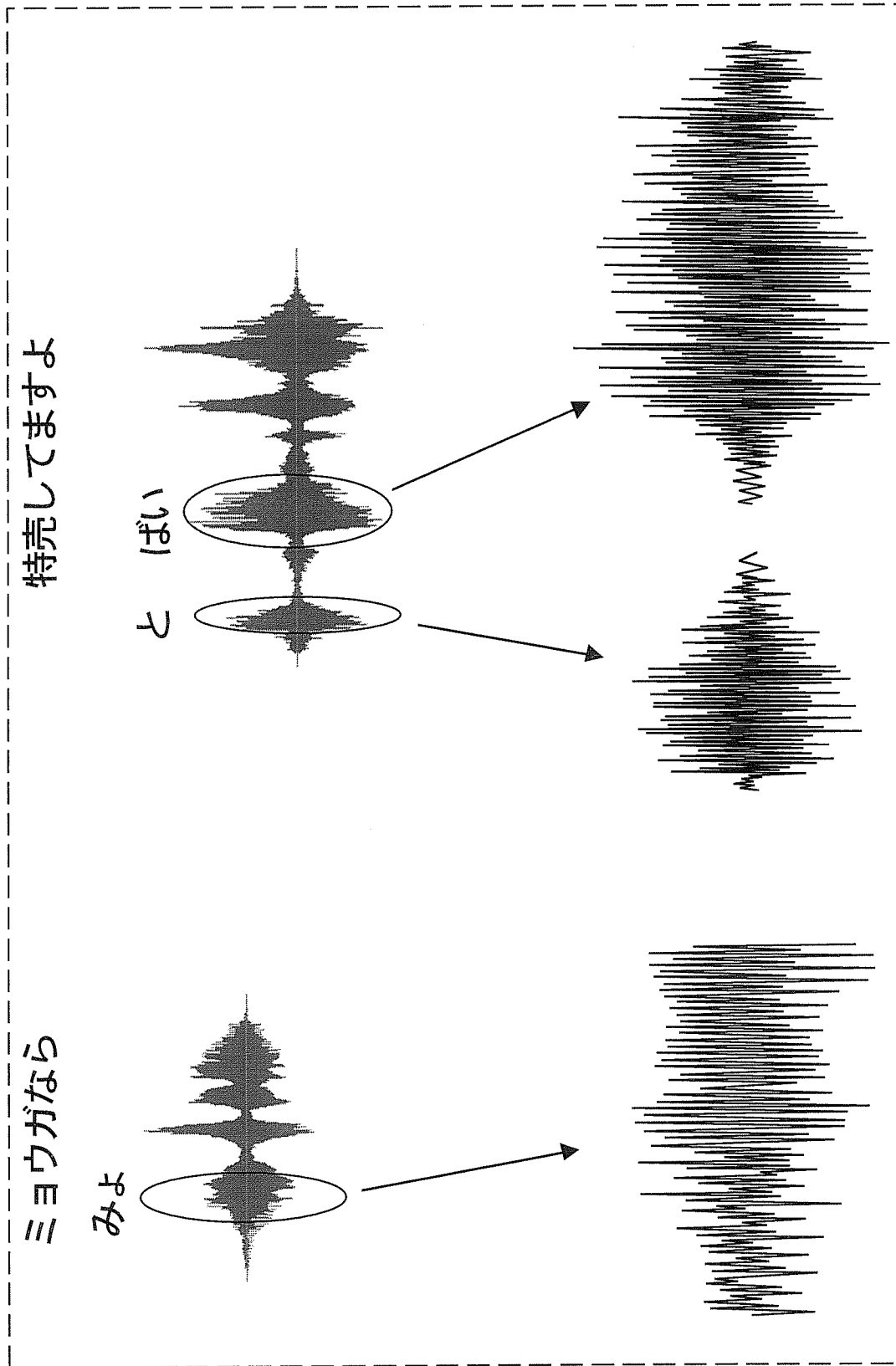
- [24] 変換の対象となる音声中の音韻を指定する力み音韻位置指定手段と、  
前記力み音韻位置指定手段により指定された音韻を表す音声波形の音源信号に  
、音韻の時間長より周期が短い周期的な振幅変動を伴う変調を施す変調手段とを備  
える

ことを特徴とする力み変換装置。

[図1]

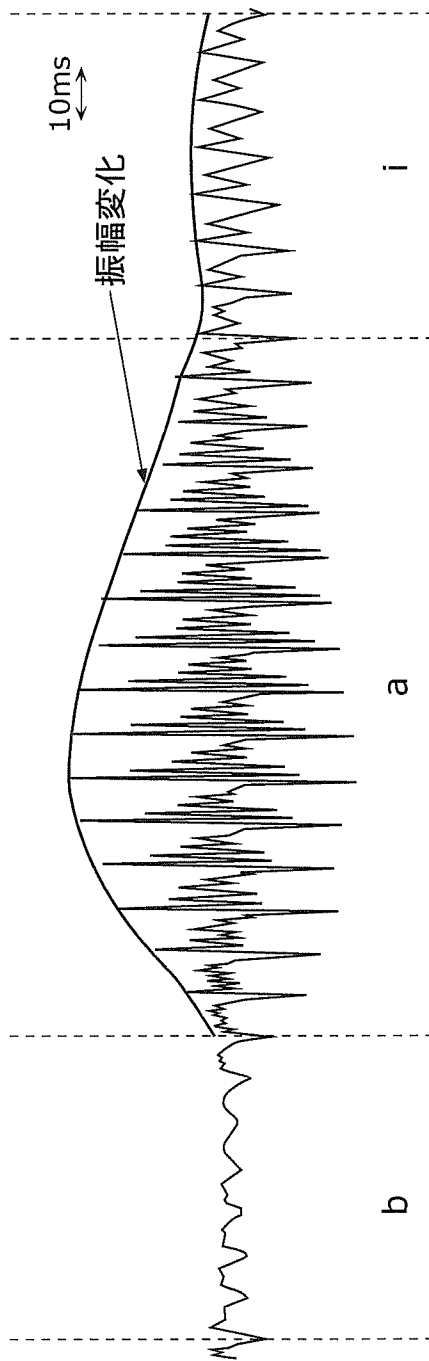


[図2]

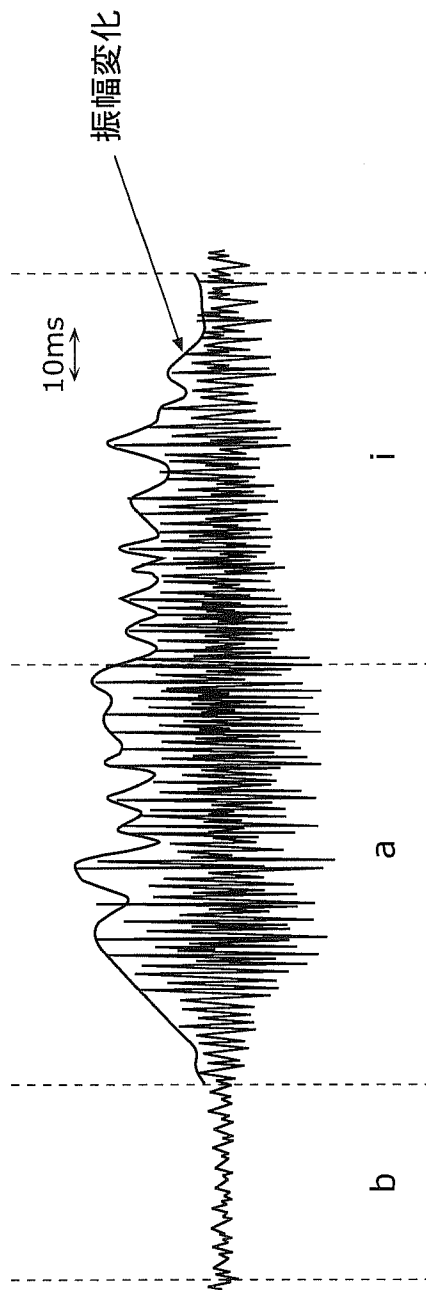




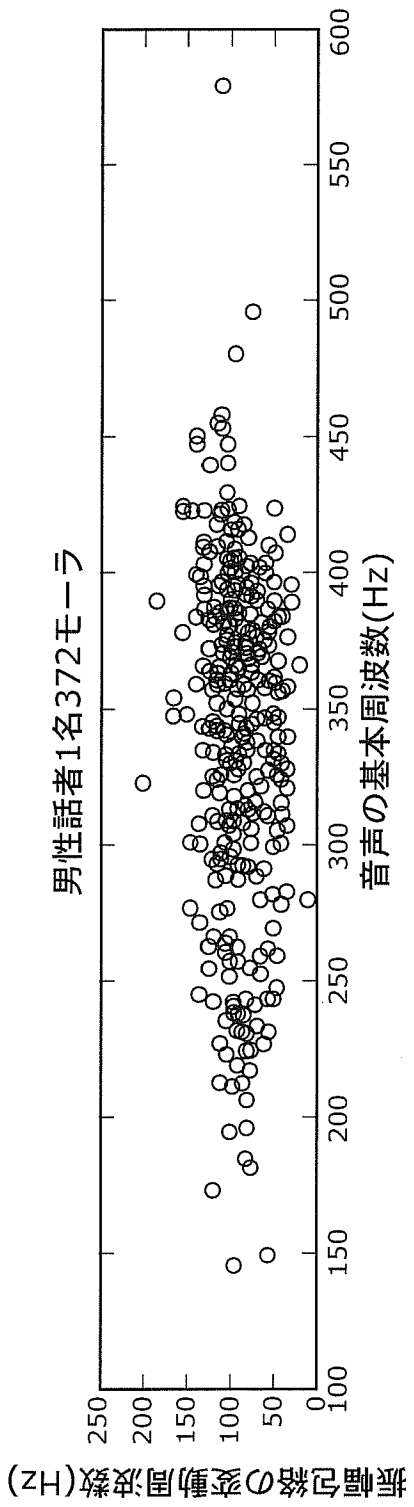
[図3A]



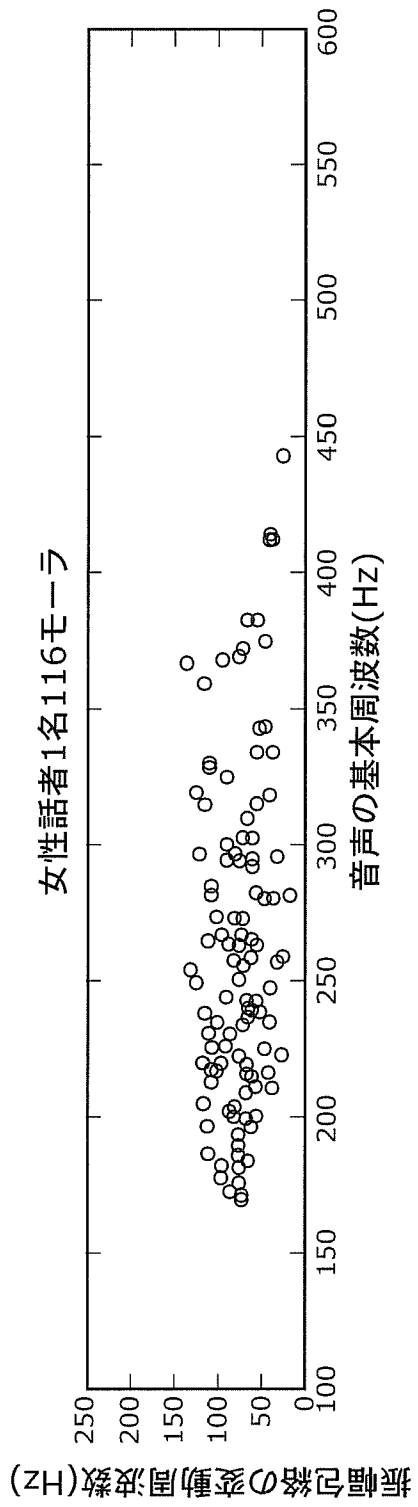
[図3B]



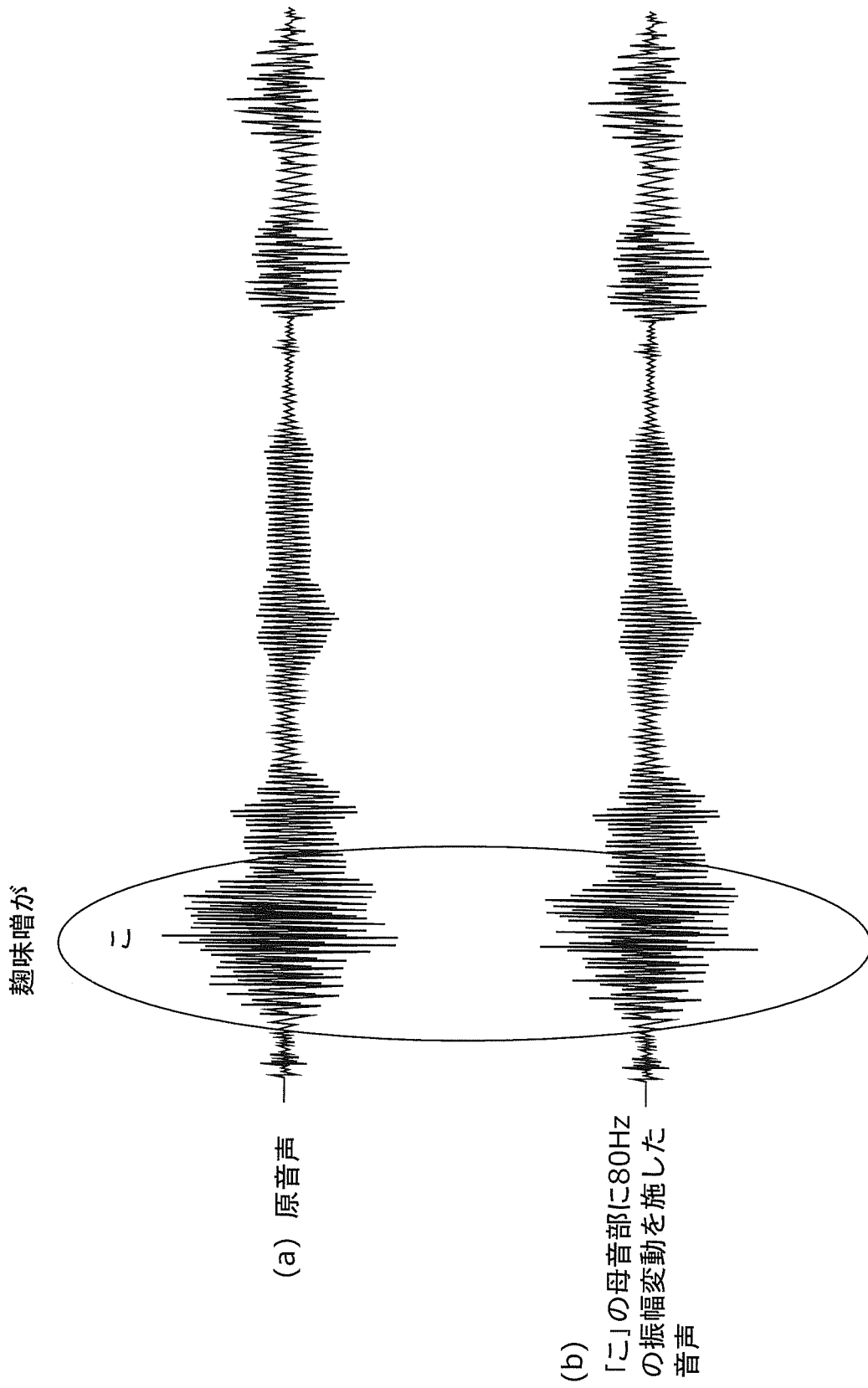
[図4A]



[図4B]



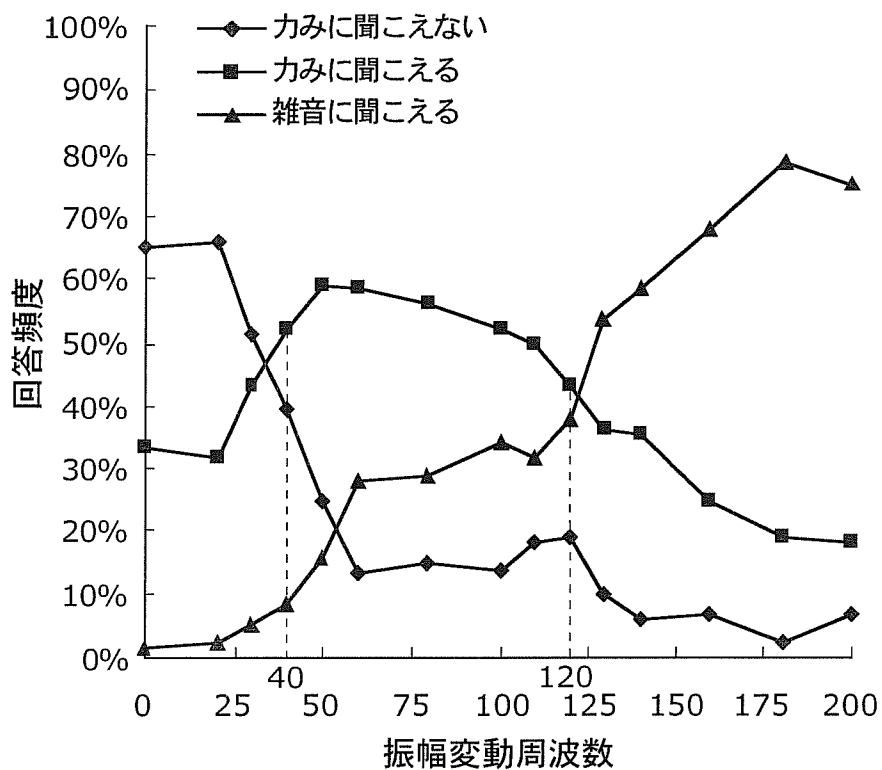
[図5]



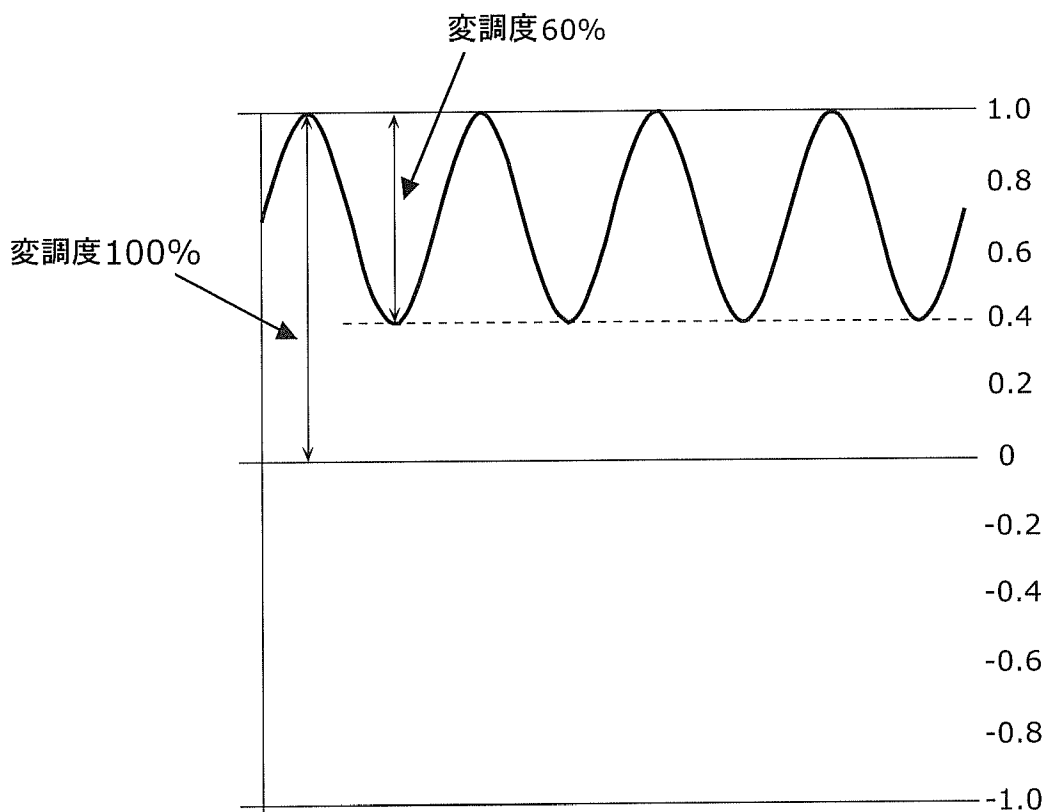
[図6]

被験者	振幅変動のある声がより「力んでいる」と判断した率 (%)
1	75
2	50
3	100
4	92
5	83
6	67
7	83
8	67
9	100
10	100
11	83
12	50
13	92
14	42
15	92
16	83
17	100
18	92
19	83
20	100

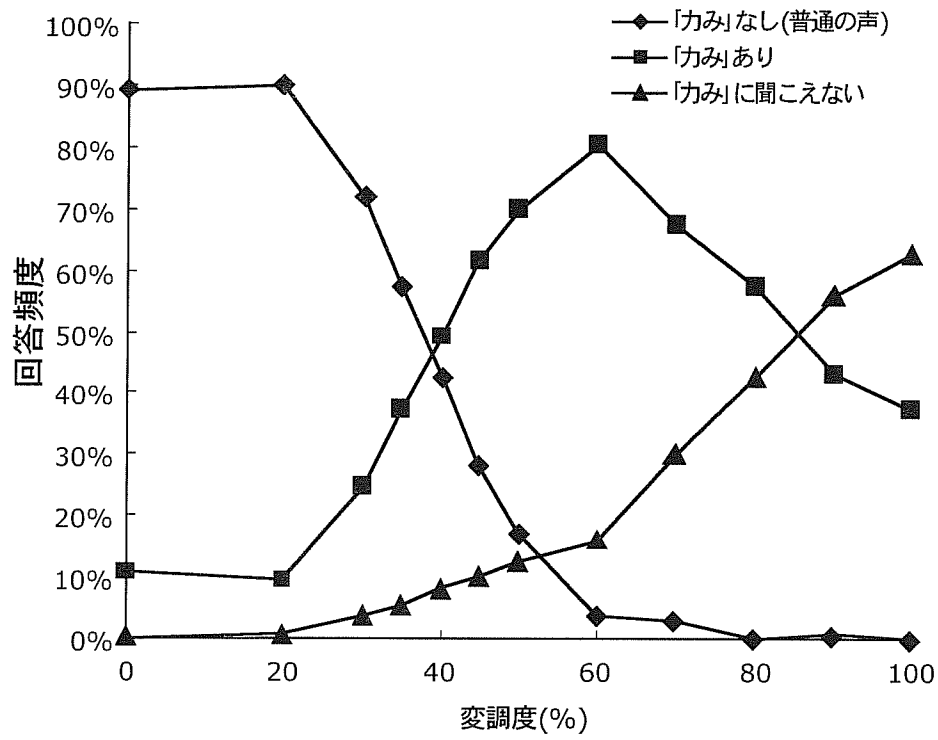
[図7]



[図8]

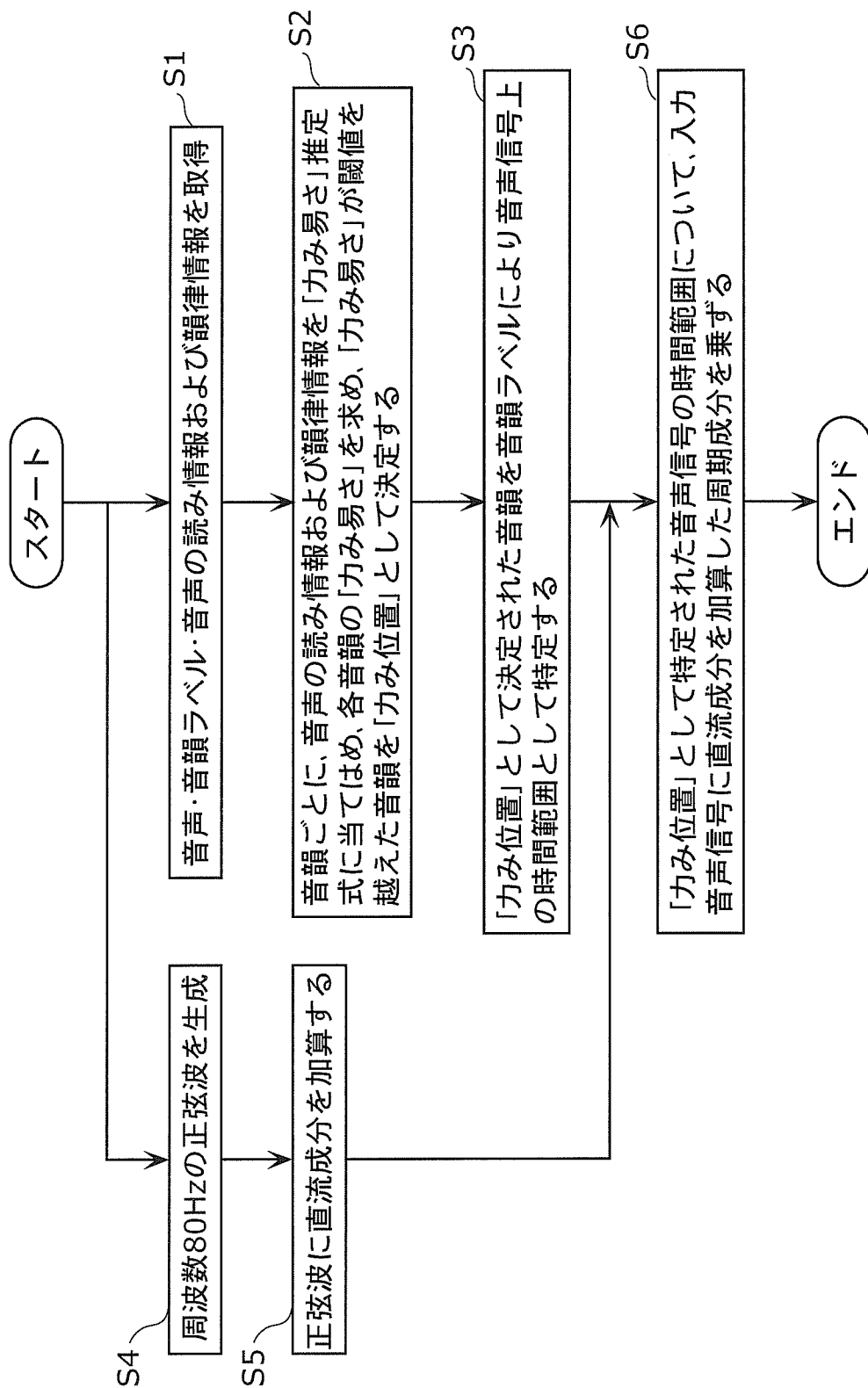


[図9]

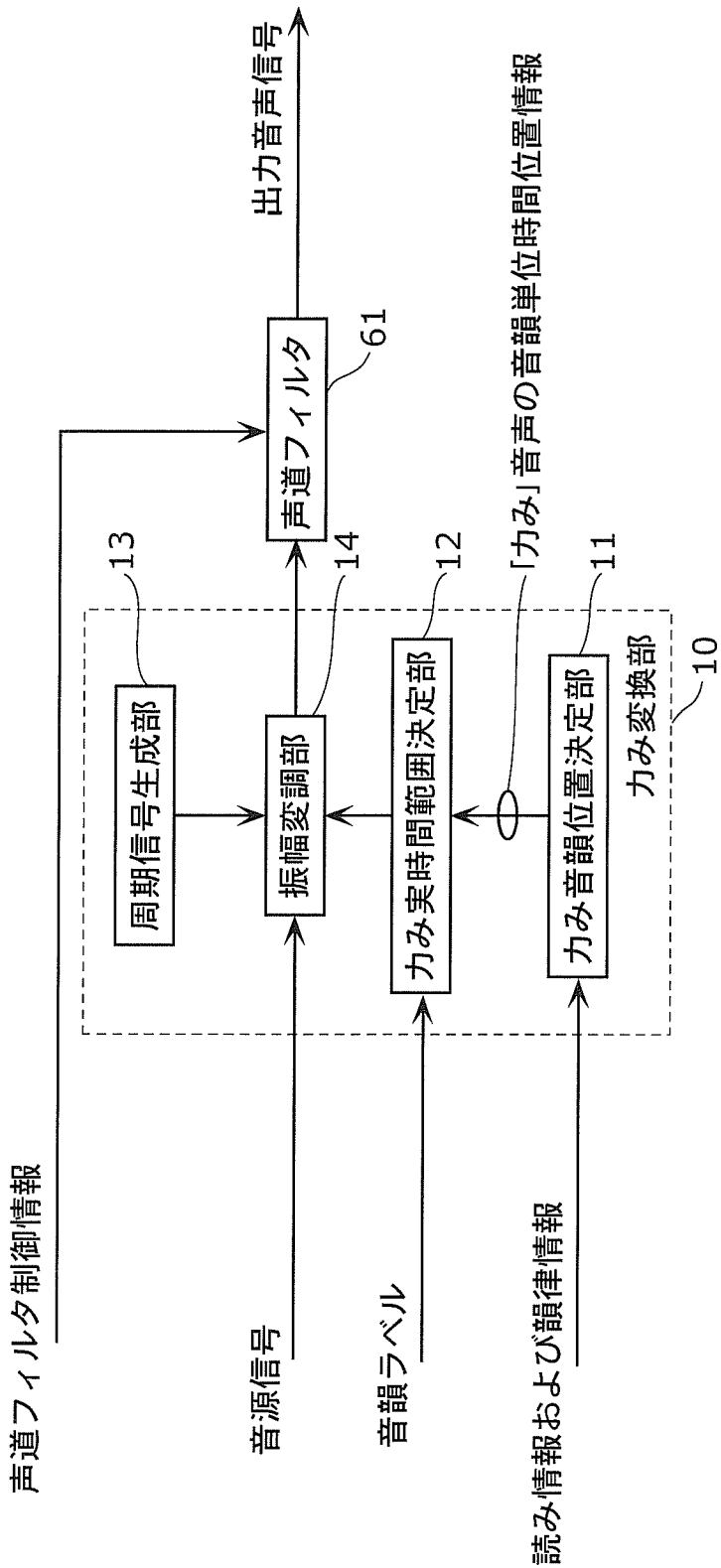




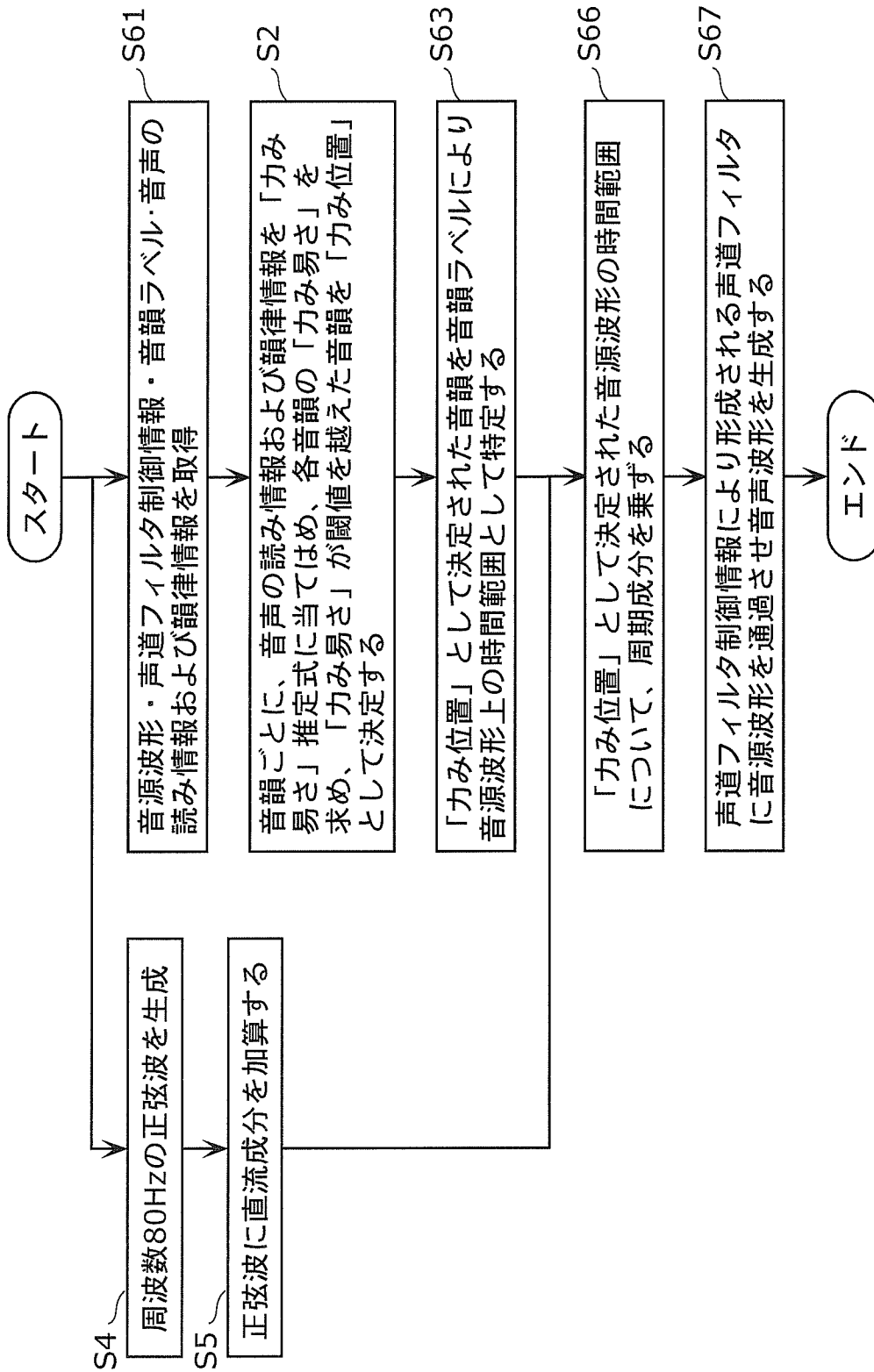
[図10]



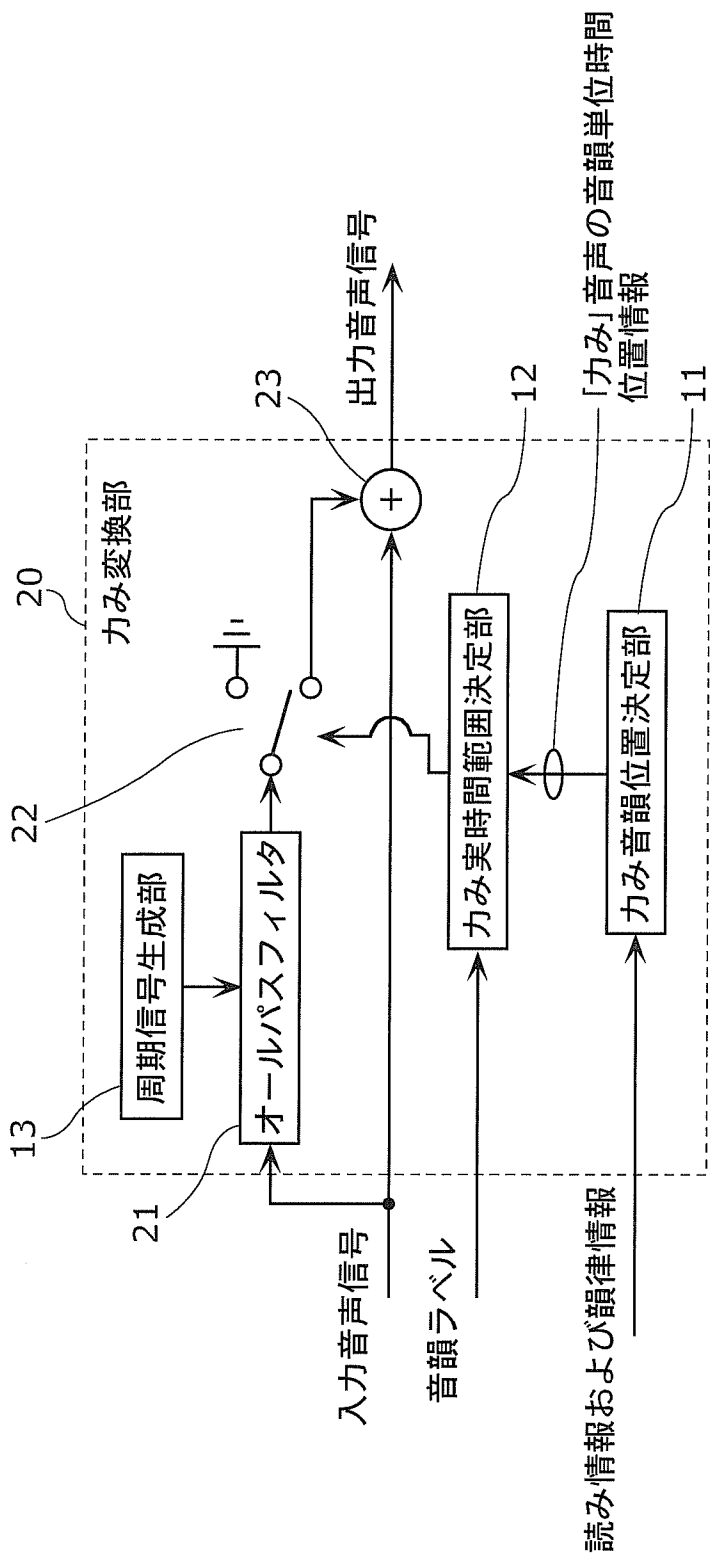
[図11]



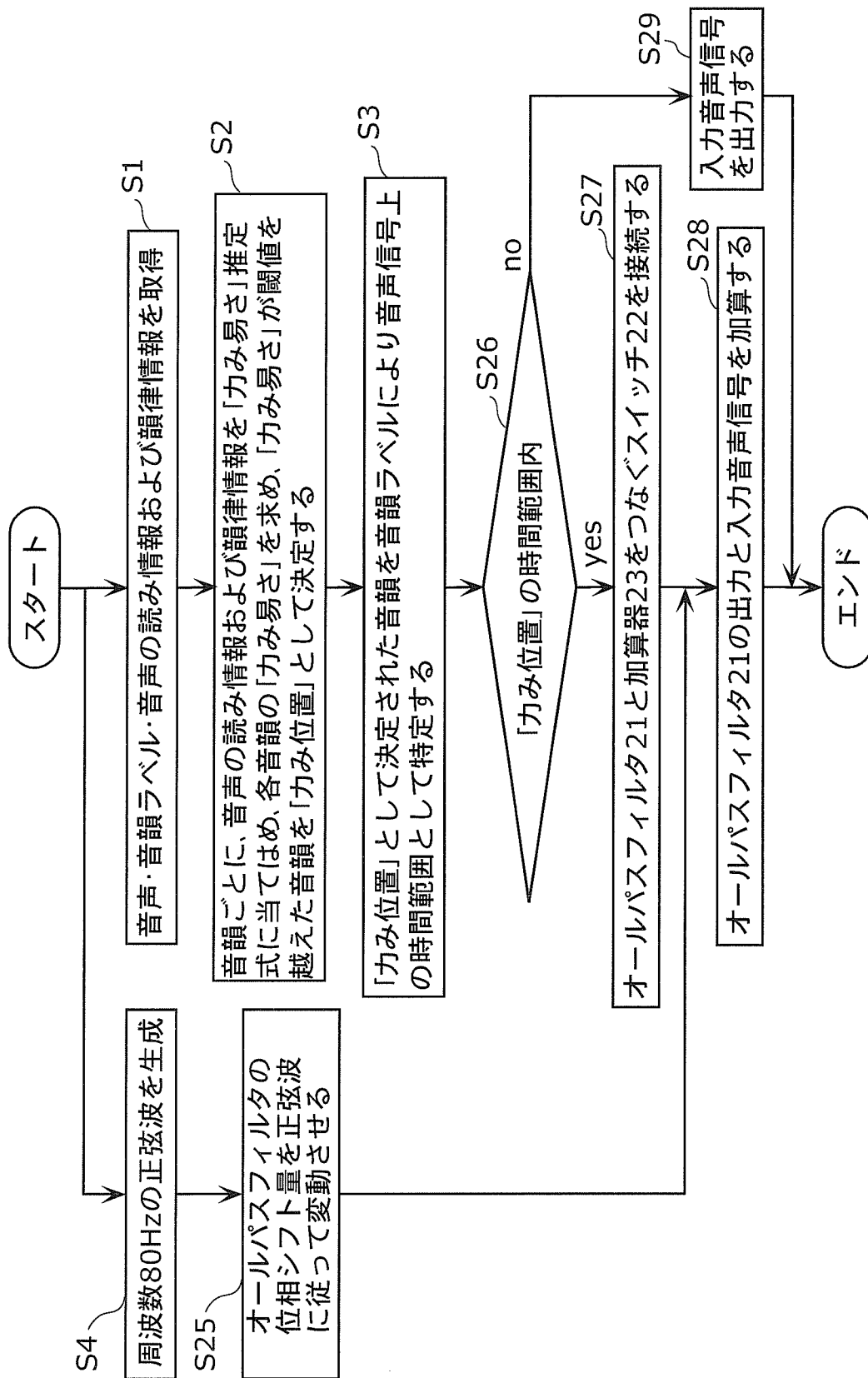
[図12]



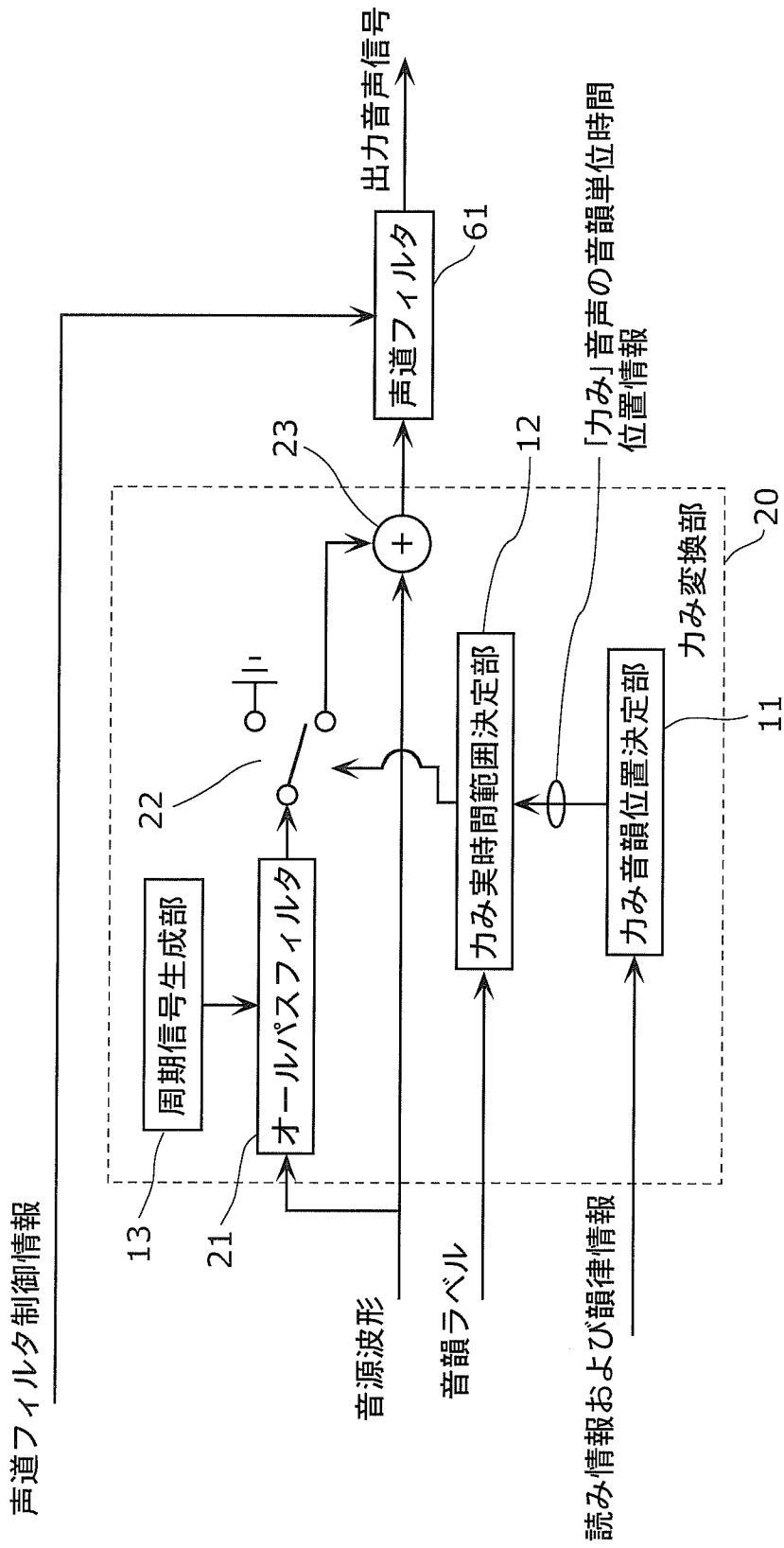
[図13]



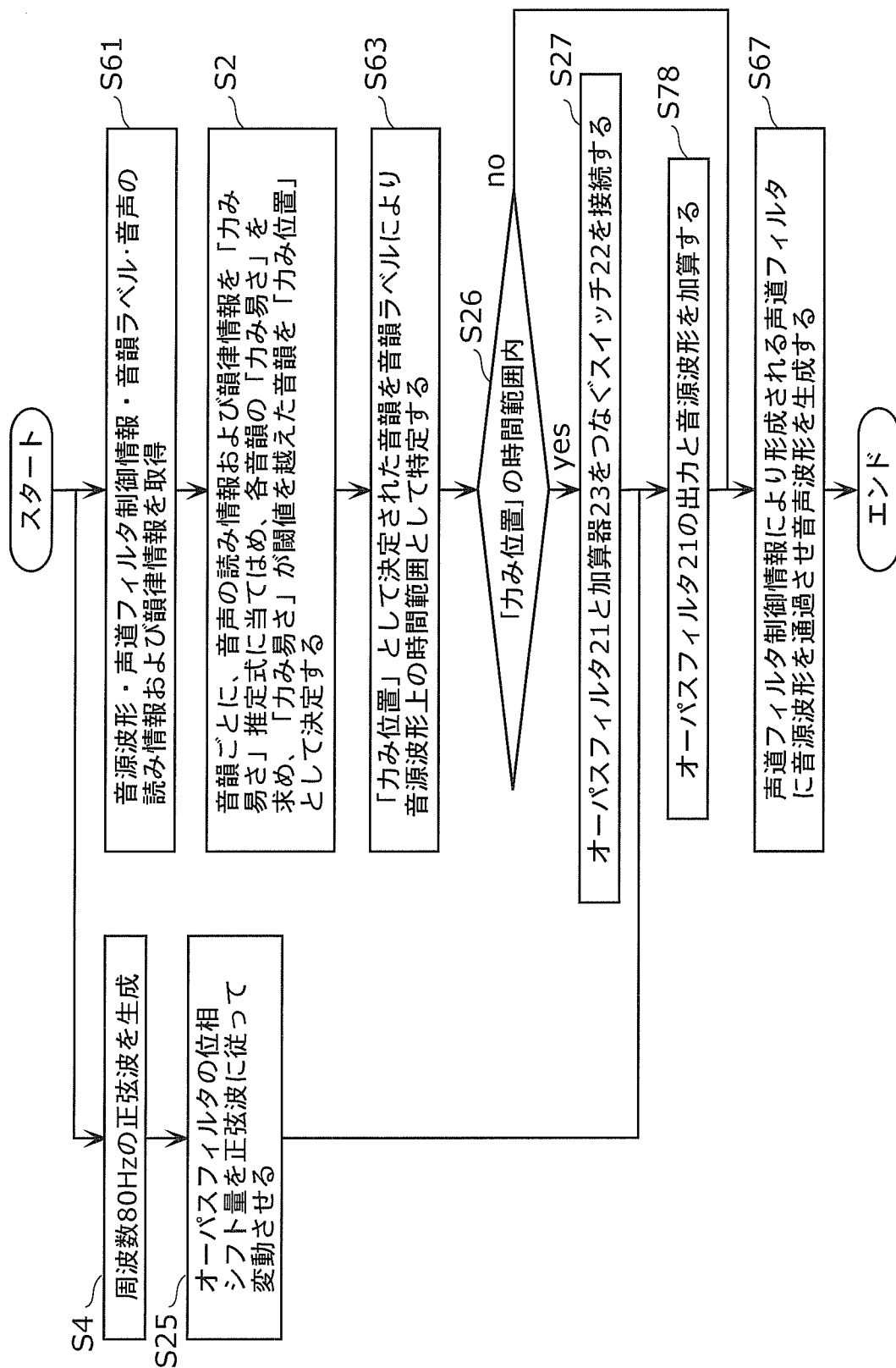
[図14]



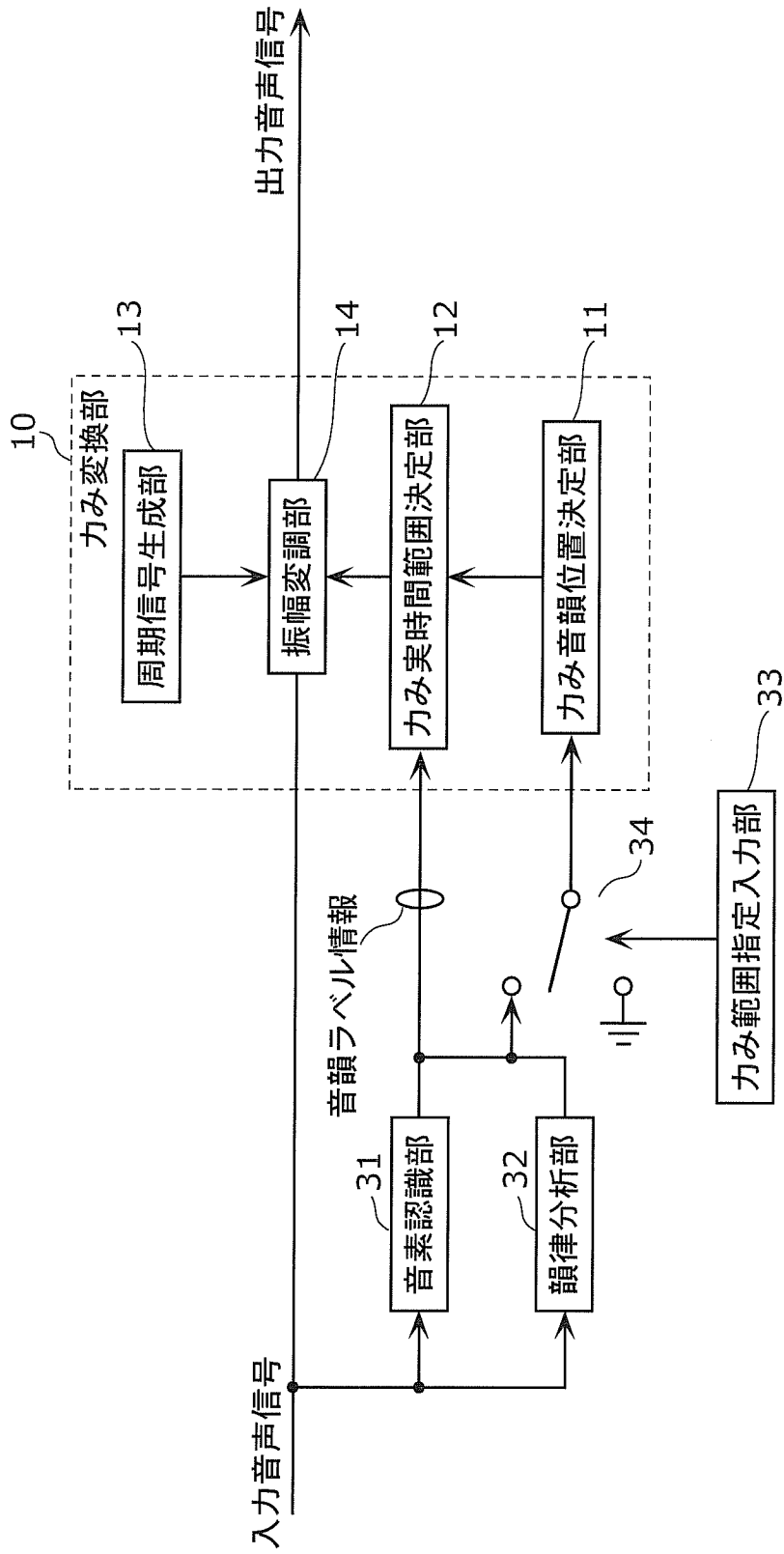
[図15]



[図16]

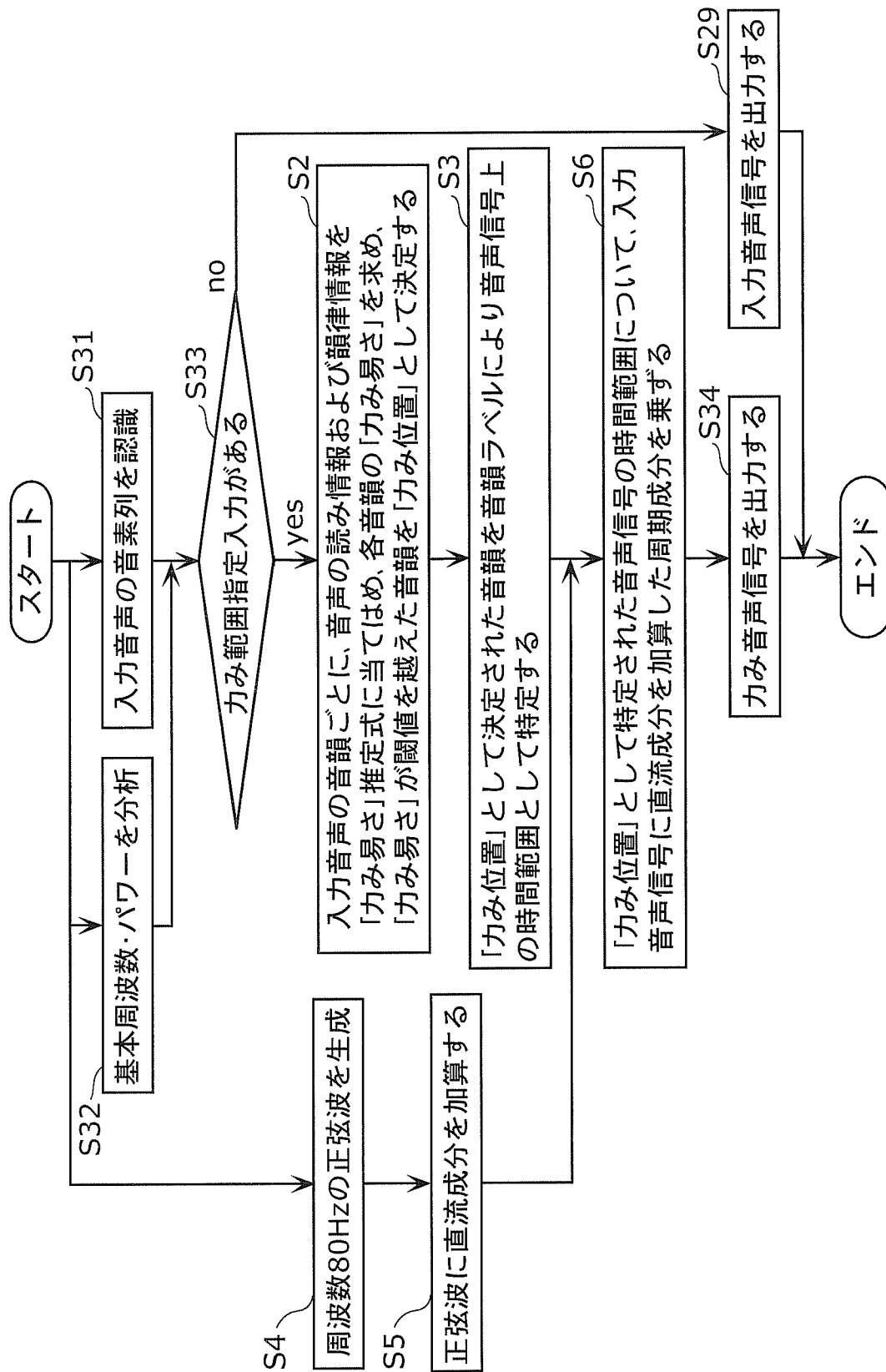


[図17]

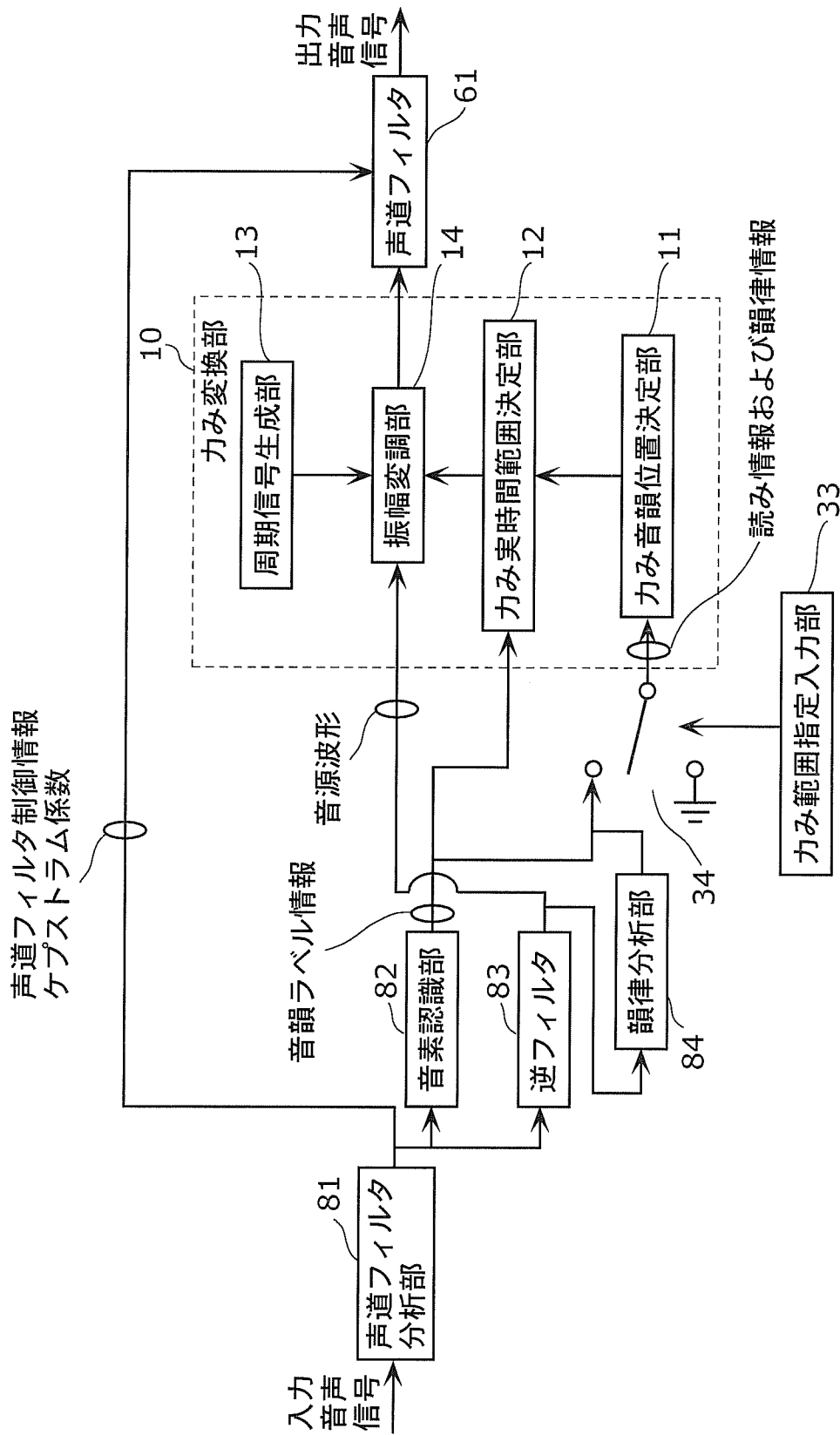




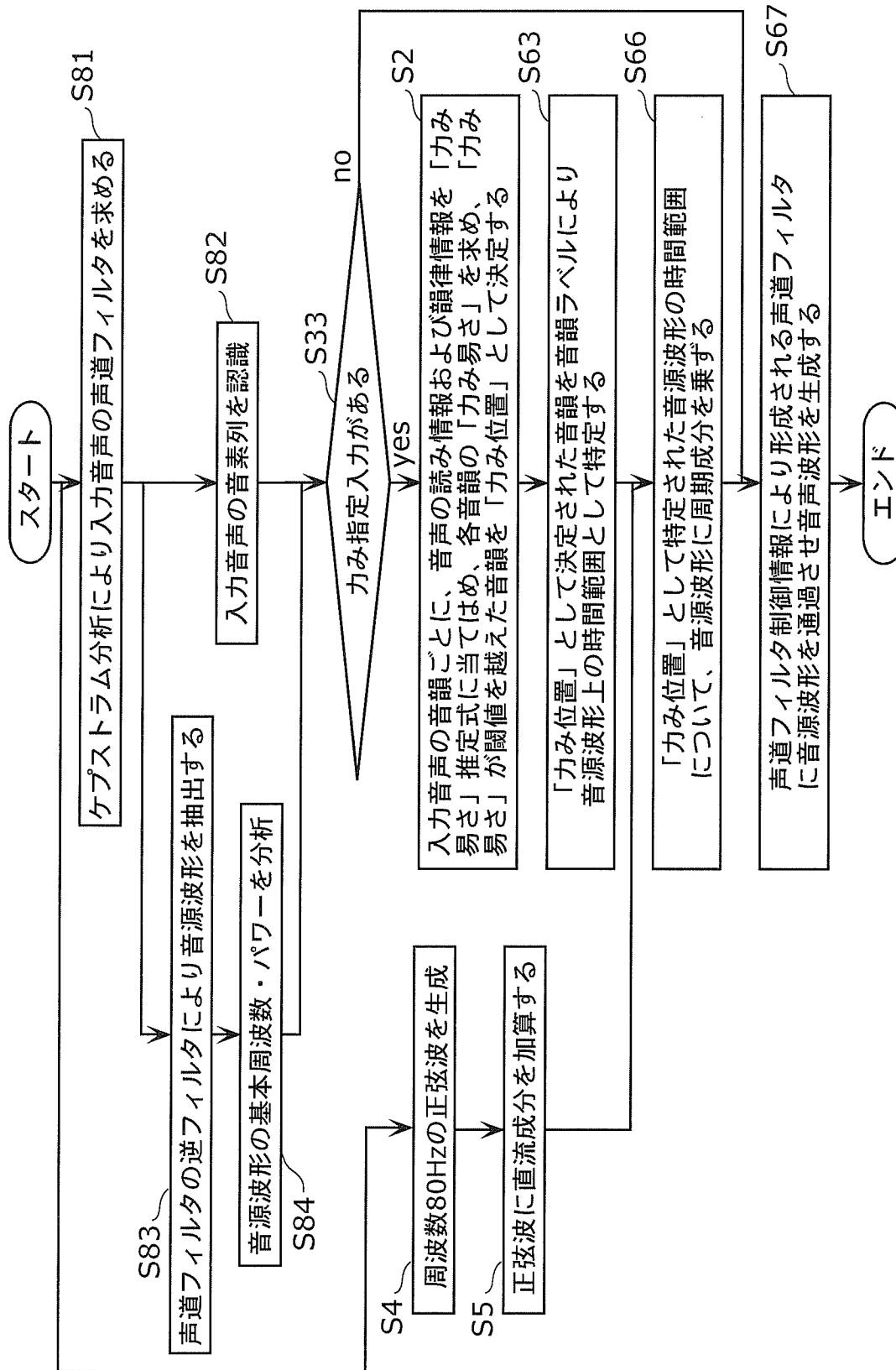
[図18]



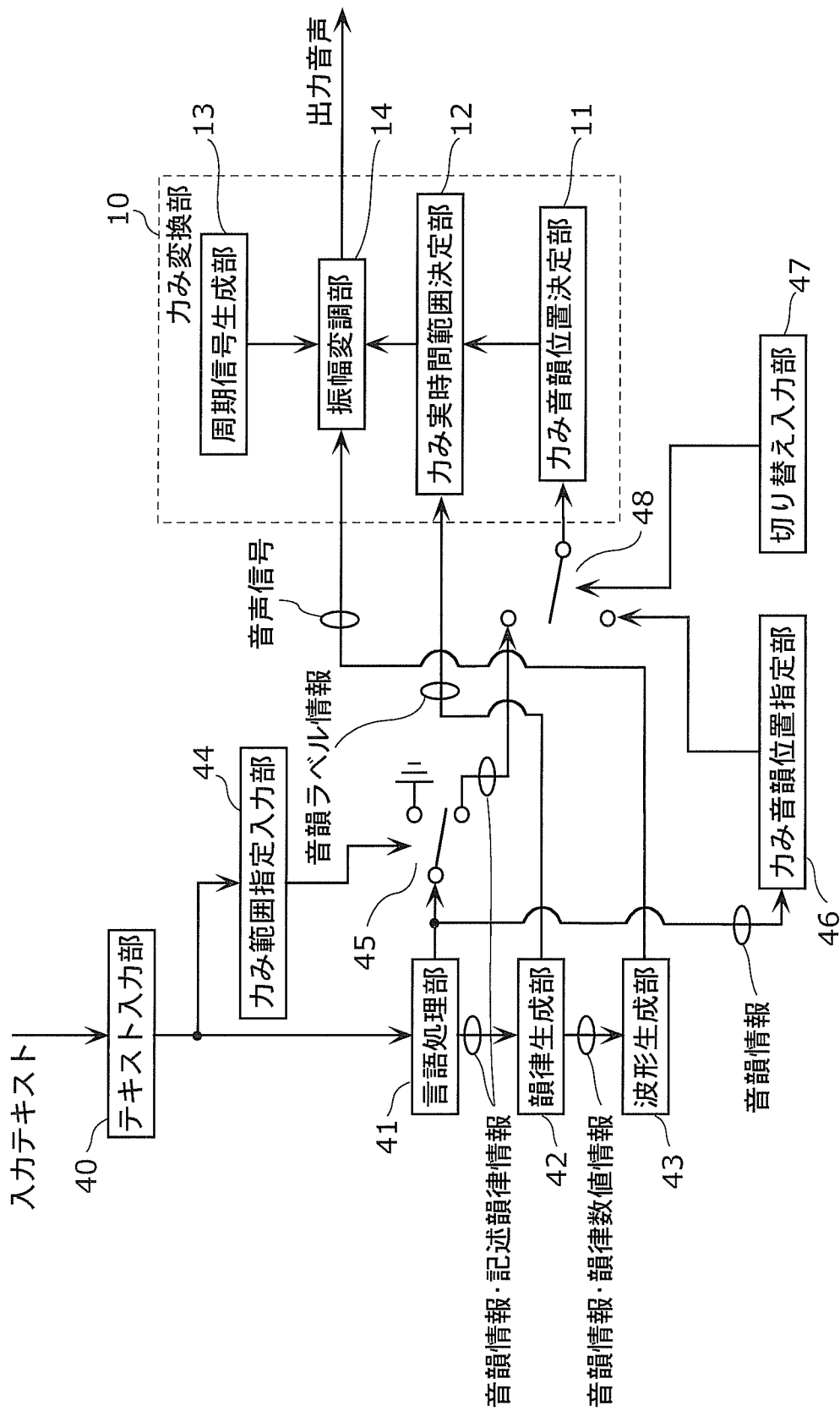
[図19]



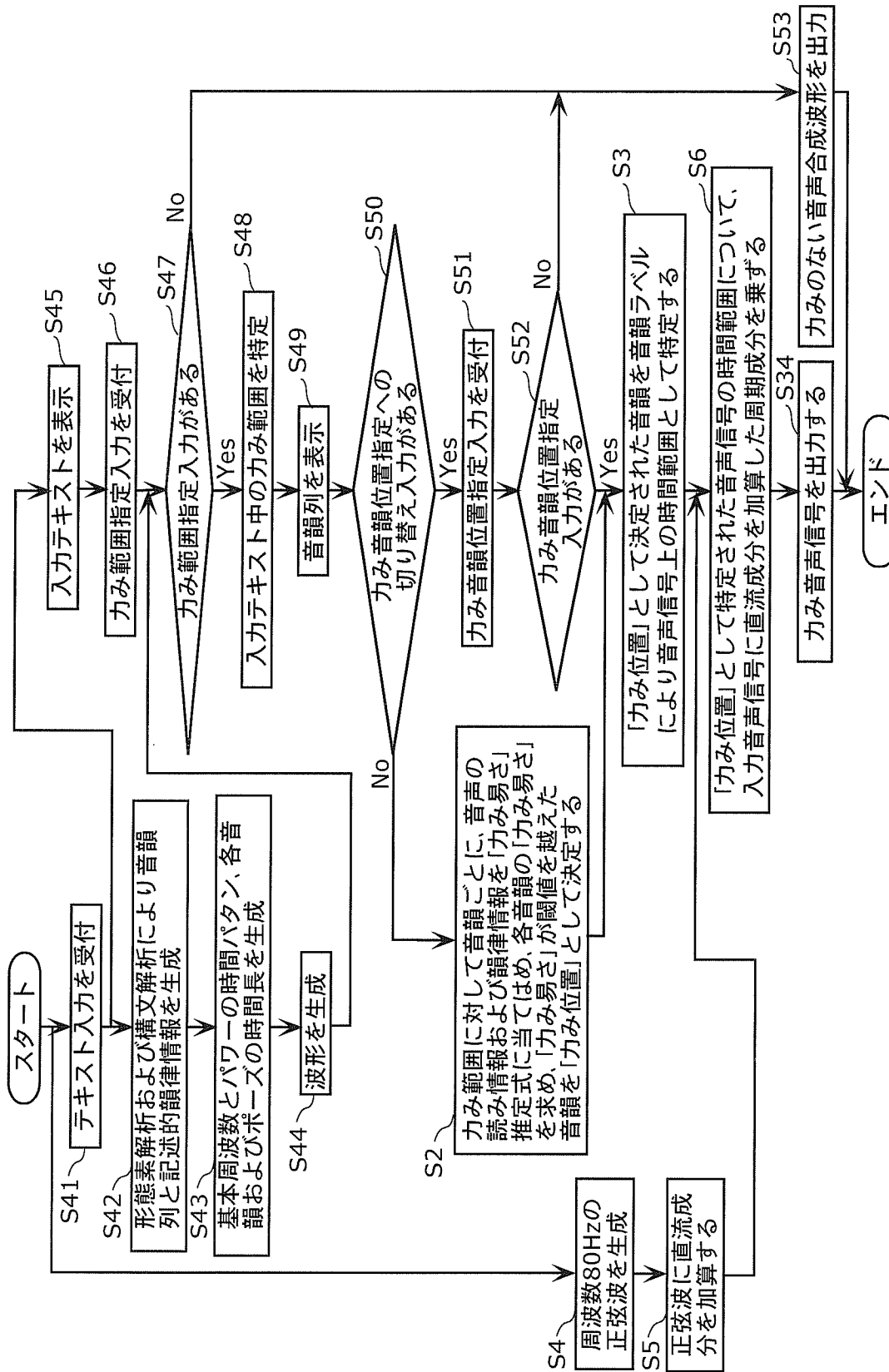
[図20]



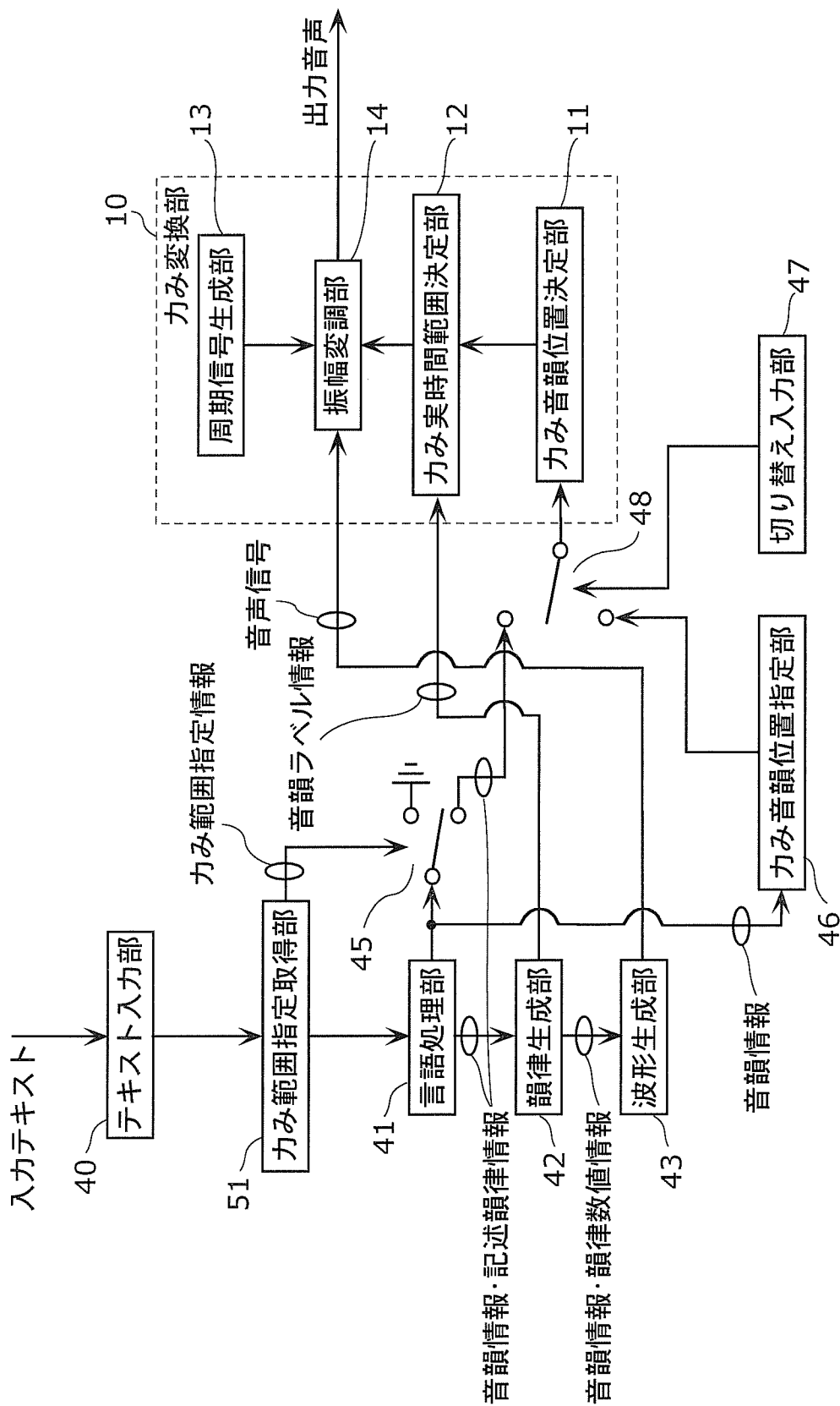
[図21]



[図22]



[図23]



[図24]

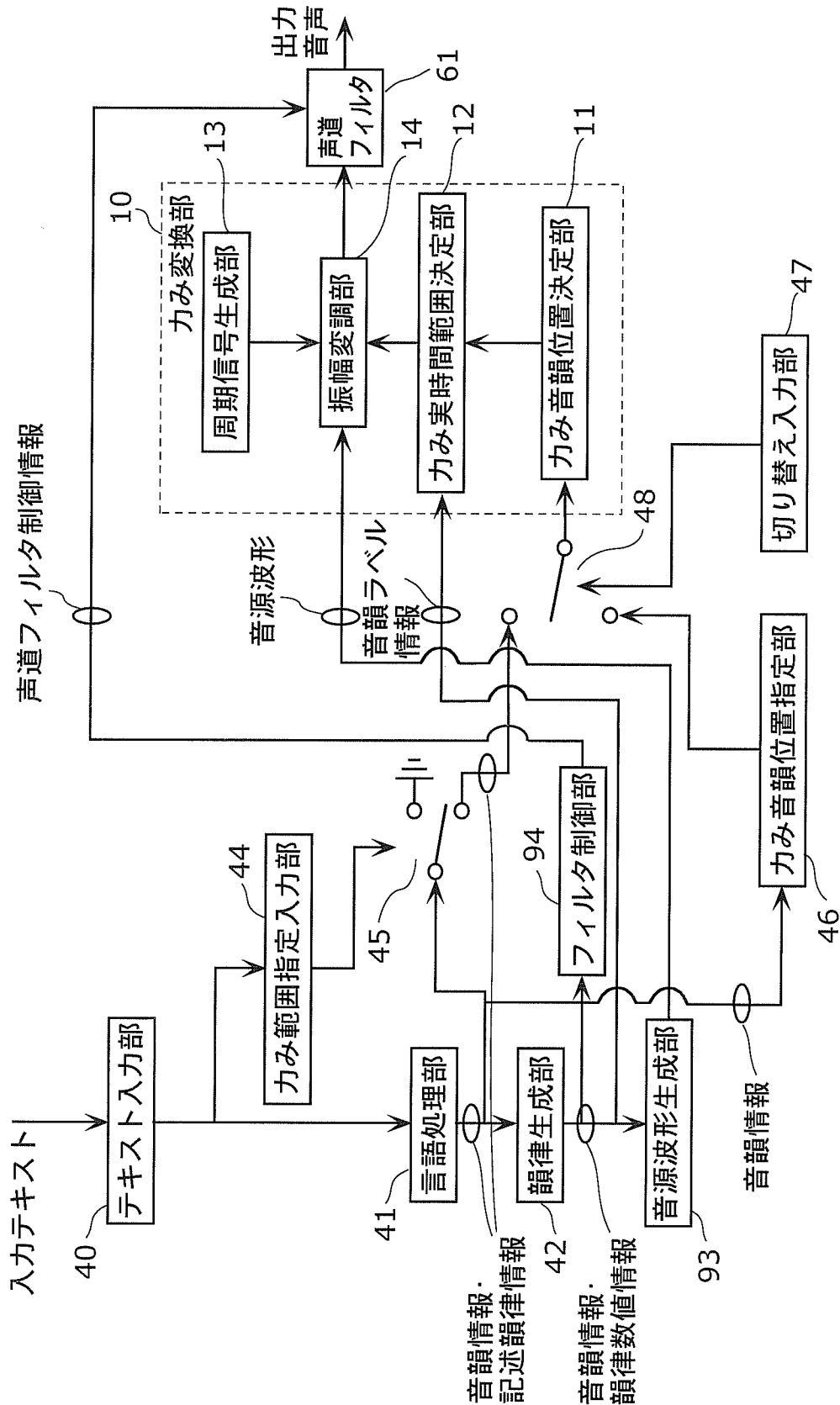
あらゆる現実をすべて自分の方へ  
 <voice quality=rikimi>捻じ曲げたのだ</voice>

[図25]

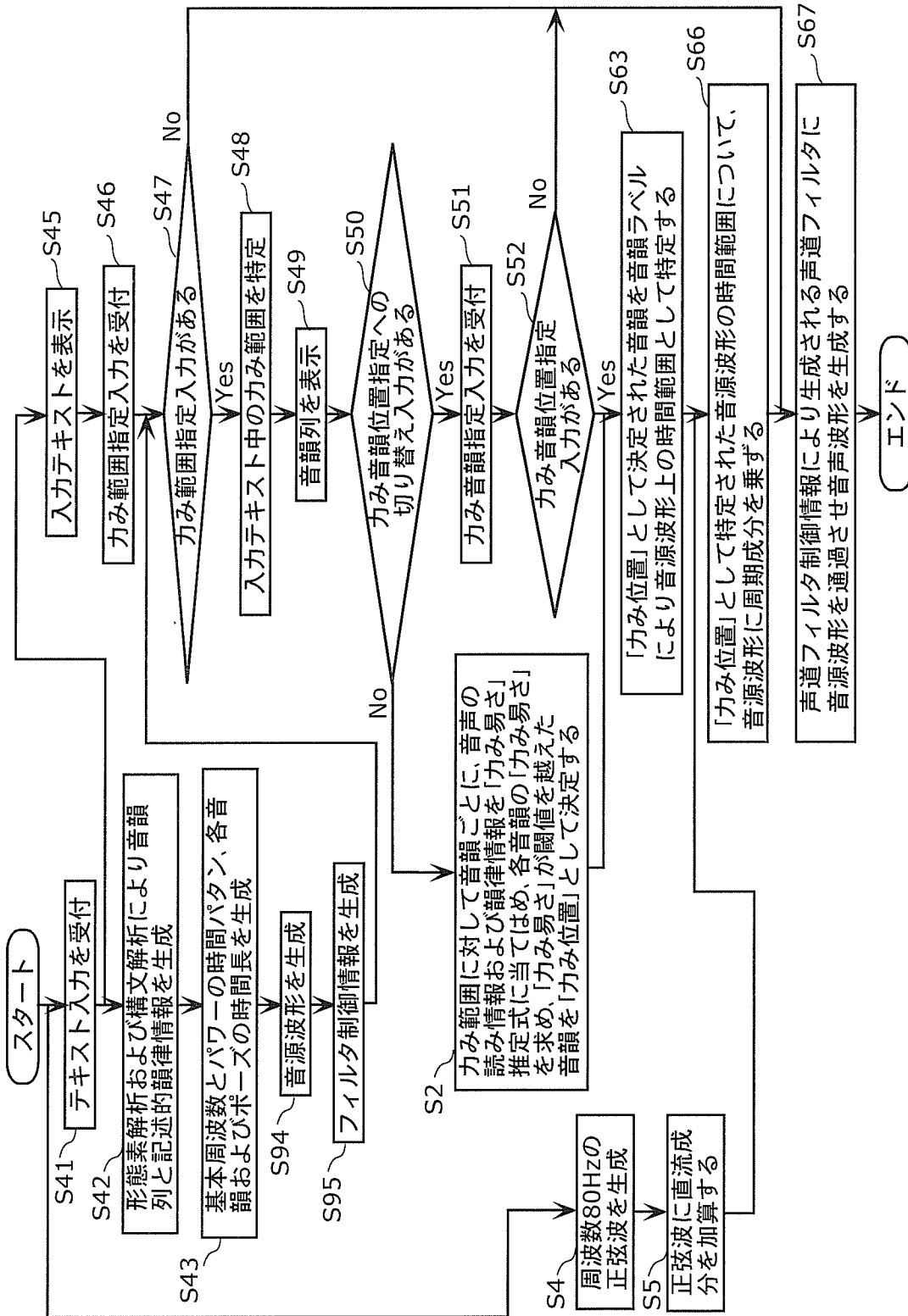
あらゆる現実をすべて

<voice quality=rikimi 5mora>自分の方へ捻じ曲げたのだ</voice>

[図26]



[図27]





## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/050815

## A. CLASSIFICATION OF SUBJECT MATTER

G10L13/08(2006.01) i, G10L21/04(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L11/00-21/06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2008

Kokai Jitsuyo Shinan Koho 1971-2008 Toroku Jitsuyo Shinan Koho 1994-2008

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JSTPlus (JDreamII)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2006/123539 A1 (Matsushita Electric Industrial Co., Ltd.), 23 November, 2006 (23.11.06), Full text; all drawings (Family: none)	1-5, 7-24
Y	JP 2002-268699 A (Sony Corp.), 20 September, 2002 (20.09.02), Full text; all drawings & US 2003/0163320 A1 & EP 1367563 A1 & WO 2002/073594 A1	1-5, 7-24
Y	JP 2002-6900 A (Kabushiki Kaisha Mega Fyujon), 11 January, 2002 (11.01.02), Full text; all drawings (Family: none)	8-11

 Further documents are listed in the continuation of Box C. See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;" document member of the same patent family

Date of the actual completion of the international search  
17 April, 2008 (17.04.08)Date of mailing of the international search report  
01 May, 2008 (01.05.08)Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/050815

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2003-84798 A (Yamaha Corp.), 19 March, 2003 (19.03.03), Full text; all drawings (Family: none)	1-24
A	JP 2002-73064 A (Yamaha Corp.), 12 March, 2002 (12.03.02), Full text; all drawings (Family: none)	1-24
A	WO 2007/010680 A1 (Matsushita Electric Industrial Co., Ltd.), 25 January, 2007 (25.01.07), Full text; all drawings (Family: none)	1-24
A	Kazuhiko MURAKAMI et al., "Onsei Gosei ni Okeru All-Pass Filter ni yoru Boon Teijobu no Yuragi Gosei", The Acoustical Society of Japan (ASJ) Heisei 5 Nen Shuki Kenkyu Happyokai Koen Ronbunshu -I-", 05 October, 1993 (05.10.93), 1-7-8, pages 607 to 608	1-24

A. 発明の属する分野の分類 (国際特許分類 (IPC))  
 Int.Cl. G10L13/08(2006.01)i, G10L21/04(2006.01)i

B. 調査を行った分野  
 調査を行った最小限資料 (国際特許分類 (IPC))  
 Int.Cl. G10L11/00-21/06

最小限資料以外の資料で調査を行った分野に含まれるもの  
 日本国実用新案公報 1922-1996年  
 日本国公開実用新案公報 1971-2008年  
 日本国実用新案登録公報 1996-2008年  
 日本国登録実用新案公報 1994-2008年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)  
 JSTPlus(JDreamII)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	WO 2006/123539 A1 (松下電器産業株式会社) 2006. 11. 23, 全文, 全図 (ファミリーなし)	1-5, 7-24
Y	JP 2002-268699 A (ソニー株式会社) 2002. 09. 20, 全文, 全図 & US 2003/0163320 A1 & EP 1367563 A1 & WO 2002/073594 A1	1-5, 7-24
Y	JP 2002-6900 A (株式会社メガフュージョン) 2002. 01. 11, 全文, 全図 (ファミリーなし)	8-11

C欄の続きにも文献が列挙されている。  パテントファミリーに関する別紙を参照。

\* 引用文献のカテゴリー  
 「A」特に関連のある文献ではなく、一般的技術水準を示すもの  
 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
 「O」口頭による開示、使用、展示等に言及する文献  
 「P」国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献  
 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの  
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの  
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの  
 「&」同一パテントファミリー文献

国際調査を完了した日 17. 04. 2008	国際調査報告の発送日 01. 05. 2008
----------------------------	----------------------------

国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 菊池 智紀 電話番号 03-3581-1101 内線 3541	5Z	3352
---	--	----	------

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 2003-84798 A (ヤマハ株式会社) 2003. 03. 19, 全文, 全図 (ファミリーなし)	1-24
A	JP 2002-73064 A (ヤマハ株式会社) 2002. 03. 12, 全文, 全図 (ファミリーなし)	1-24
A	WO 2007/010680 A1 (松下電器産業株式会社) 2007. 01. 25, 全文, 全図 (ファミリーなし)	1-24
A	村上和彦 他, "音声合成におけるオールパスフィルタによる母音定常部の 揺らぎ合成", 日本音響学会平成5年秋季研究発表会講演論文集-I-, 1993. 10. 05, 1-7-8, p. 607-608	1-24