

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7295839号  
(P7295839)

(45)発行日 令和5年6月21日(2023.6.21)

(24)登録日 令和5年6月13日(2023.6.13)

(51)国際特許分類 F I  
 G 1 0 L 15/18 (2013.01) G 1 0 L 15/18 3 0 0 H  
 G 1 0 L 15/10 (2006.01) G 1 0 L 15/10 3 0 0 F  
 G 1 0 L 15/197(2013.01) G 1 0 L 15/197

請求項の数 30 (全30頁)

(21)出願番号	特願2020-501252(P2020-501252)	(73)特許権者	511114818 エスシーティアイ ホールディングス、 インク アメリカ合衆国 ワシントン ディーシー 2 0 0 1 5 ピーオーボックス 4 2 5 9 8
(86)(22)出願日	平成30年7月10日(2018.7.10)	(74)代理人	100073184 弁理士 柳田 征史
(65)公表番号	特表2020-527253(P2020-527253 A)	(74)代理人	100123652 弁理士 坂野 博行
(43)公表日	令和2年9月3日(2020.9.3)	(74)代理人	100175042 弁理士 高橋 秀明
(86)国際出願番号	PCT/US2018/041395	(72)発明者	ビンソン, ダレル ティー アメリカ合衆国 メリーランド州 2 0 9 1 0 シルヴァー スプリング フェントン ストリート 8 6 3 0 ナンバー 5 2 0 最終頁に続く
(87)国際公開番号	WO2019/014183		
(87)国際公開日	平成31年1月17日(2019.1.17)		
審査請求日	令和3年7月6日(2021.7.6)		
(31)優先権主張番号	62/530,802		
(32)優先日	平成29年7月10日(2017.7.10)		
(33)優先権主張国・地域又は機関	米国(US)		

(54)【発明の名称】 音節に基づく自動音声認識

(57)【特許請求の範囲】

【請求項1】

データ処理方法において、

特定の言語で話された単語から生成された生成記号ストリームを、音響処理システムから計算システムで受信する工程と、

複数の生成パターンを、前記生成記号ストリームから抽出する工程と、

生成パターンから標準パターンへの1つ以上のマッピングについての条件付き確率を含むものである、保存された生成 - 標準マッピングデータを用いて、候補音節、および、各前記候補音節の確率を、前記複数の生成パターンから生成する工程と、

1つ以上の候補音節から綴りパターンへのマッピングについての条件付き確率を含むものである、保存された音節 - 綴りパターンマッピングデータを用いて、候補綴りパターン、および、各前記候補綴りパターンの確率を、前記候補音節から生成する工程と、

前記生成記号ストリームの綴り表記を、少なくとも部分的には各前記候補綴りパターンの前記確率に基づいて生成する工程とを含む方法。

【請求項2】

前記生成記号ストリームは、母音間子音 ( I V C ) および母音近傍 ( V N ) 単位を含む音素配列単位へと分割されるものであり、その分割は、該生成記号ストリームの各記号について、

3 記号バッファをゼロに、I V C 累積バッファをゼロとなるように初期化して、生成

10

20

記号を順に前記 3 記号バッファに追加する工程と、

記号を前記 3 記号バッファに追加した後に、該 3 記号バッファの中央の記号が母音であり、したがって、それらの 3 つの記号は、VN を含むかを決定して、前記 VN を保存する工程と、

追加された記号が子音の場合は、前記子音を、前記 IVC 累積バッファに加える工程と、

次に追加された記号が子音ではない場合は、前記記号を前記 IVC 累積バッファに、IVC として保存する工程と、

前記生成記号ストリームが続く限りは処理を継続する工程と

を順に行うことによって行われるものである、請求項 1 に記載のデータ処理方法。

【請求項 3】

前記生成記号ストリームは、母音間子音 (IVC) 単位および母音近傍 (VN) 単位を含む音素配列単位に分割された生成音素配列ストリームであり、前記候補音節を生成する工程は、前記生成音素配列ストリームを、標準音素配列ストリームにマッピングする工程を含み、その工程は、

処理待ち行列を、空となるように初期化する工程と、

生成 IVC と生成 VN の交互のストリームを受信する工程と、

アラインメントされた生成 IVC および標準 IVC の予め計算保存されたマッピングデータを用いる工程であって、生成 IVC を 1 つ以上の標準 IVC に、保存され関連付けられた条件付き確率と共にマッピングするものである工程と、

アラインメントされた生成および標準 VN の保存されたマッピングデータを用いる工程であって、生成 VN を 1 つ以上の標準 VN に、保存され関連付けられた条件付き確率と共にマッピングするものである工程と

によって行われるものである、請求項 1 に記載のデータ処理方法。

【請求項 4】

前記マッピングされた標準音素配列ストリームを更に処理する工程を、含み、その工程は、

各前記標準 IVC が、直前の標準 VN パターンの組の少なくとも 1 つの要素から整合して連続したものであるかを決定し、そのような標準 VN パターンの少なくとも 1 つとも整合しない標準 IVC を除外する工程と、

各前記標準 VN が、直前の標準 IVC パターンの組の少なくとも 1 つの要素から整合して連続したものであるかを決定し、そのような標準 IVC パターンの少なくとも 1 つとも整合しない標準 VN を除外する工程と、

整合しない IVC および VN を除外した後に、条件付き確率を再び正規化する工程とによって行われるものである、請求項 3 に記載のデータ処理方法。

【請求項 5】

前記候補音節を生成する工程は、

整合した標準 IVC パターンを、音節のコーダ オンセットシーケンスへと分割する工程と、

分割された整合した標準 IVC パターンのコーダ部分を用いて、残りの IVC パターンおよび VN パターンから増やされたオンセットを通る整合した経路を完成する工程であり、各前記整合した経路は候補音節を形成するものである工程と、

保存された音節表にない候補音節を除外する工程と、

各前記候補音節の確率を計算する工程と、

前記候補音節を保存する工程と

を含むものである、請求項 1 に記載のデータ処理方法。

【請求項 6】

各前記候補音節について確率を生成する工程は、

特定の生成記号の組の各候補音節について、

前記候補音節についての前記 VN の確率を、前記生成 - 標準マッピングデータの保存計算された確率から特定する工程、

10

20

30

40

50

前記候補音節のオンセットの確率を、前記オンセットがそれに基づきうる各標準 IVC についての前記生成 - 標準マッピングデータの保存計算された確率に基づいて特定する工程と、

前記候補音節のコーダの確率を、前記コーダがそれに基づきうる各標準 IVC についての前記生成 - 標準マッピングデータの保存計算された確率に基づいて特定する工程と、

前記 VN の前記確率、前記オンセットの前記確率、および、前記コーダの前記確率を集計して、音節確率値とする工程と、

前記特定の生成記号の組について、前記候補音節の前記音節確率値の合計を正規化する工程と

を含むものである、請求項 5 に記載のデータ処理方法。

10

【請求項 7】

候補音節を綴りパターンにマッピングする工程は、

アラインメントされた標準音節分けした発音および綴り音節パターンを含むデータを保存する工程と、

待ち行列を、空となるように初期化して、前記候補音節を順に提供し、学習保存された音節分けした綴りパターンの組にマッピングする工程と、

前記パターンと関連付けられた単語位置コードを用いて、無効パターンを除外する工程と、

前記マッピングされた綴りパターンを保存する工程と

を含むものである、請求項 1 に記載のデータ処理方法。

20

【請求項 8】

音節位置制約、および、統計モデルを適用して、出力綴りパターンを選択する工程を、更に含み、その工程は、

綴りパターンを、待ち行列を通して、重なる音節のシーケンスが順次処理される複数の音節サイズの単位のスライディングウインドウの形態で順に処理する工程と、

綴りパターンの組を受信した時に、各綴りパターンの組に対応する n グラムモデルに保存されたデータを参照する工程と、

n グラムモデルの条件付き確率を合計して、経路確率を生成する工程と、

前記経路確率を保存する工程と

によって行われるものである、請求項 1 に記載のデータ処理方法。

30

【請求項 9】

アラインメントされた生成パターンおよび標準パターンを含むデータを保存する工程と、

前記保存されたマッピングデータの生成パターンおよび標準パターンの各組合せについて、前記標準パターンの確率を、該保存されたマッピングデータの前記組合せの発生回数に対応する値と、異なる生成パターンの前記標準パターンの発生回数に対応する値との関数として、計算する工程と、

前記生成パターンおよび前記標準パターンの組合せを、対応する前記生成 - 標準マッピングデータの前記組合せの計算した確率と共に保存する工程と

を更に含む、請求項 1 に記載のデータ処理方法。

【請求項 10】

40

アラインメントされた音節分け標準発音および綴りパターンを含むデータを保存する工程と、

前記保存されたマッピングデータの音節分け標準発音および綴りパターンの各組合せについて、前記パターンの確率を、該保存されたマッピングデータの前記組合せの発生回数に対応する値と、該保存されたマッピングデータの前記音節分け標準発音の発生回数に対応する値との関数として、計算する工程と、

前記音節分け標準発音および綴りパターンの組合せを、対応する前記生成 - 標準マッピングデータの前記組合せの計算した確率と共に保存する工程と

を更に含む、請求項 1 に記載のデータ処理方法。

【請求項 11】

50

文字起こしされた特定の言語で話されたか、または、書かれた音声を含む総語数コーパスを保存する工程と、

特定の言語の単語の音節分けした発音および対応する綴りパターン含む発音辞書を保存する工程と、

前記発音辞書を用いて、前記総語数コーパスを、音節分けした綴りのチャンクコーパスに変換する工程と、

前記音節分けした綴りのチャンクコーパスを用いて、各  $n$  グラムパターンについて関連回数を累積して、各候補綴り音節パターンについて確率を生成する工程と

を更に含む、請求項 1 に記載のデータ処理方法。

【請求項 1 2】

文字起こしされた特定の言語で話されたか、または、書かれた音声を含む総語数コーパスを保存する工程と、

特定の言語の単語の音節分けした発音および対応する綴りパターン含む発音辞書を保存する工程と、

前記発音辞書を用いて、前記総語数コーパスを、発音音節コーパスに変換する工程と、

前記発音音節コーパスを用いて、各  $n$  グラムパターンについて関連回数を累積して、各候補発音音節パターンについて確率を生成する工程と

を更に含む、請求項 1 に記載のデータ処理方法。

【請求項 1 3】

訓練中に、与えられた候補綴りパターンが、対象綴りパターンに一致しない最も尤度が高い綴りパターンであり、

与えられた候補綴りパターンを含む多音節シーケンスの拡張コンテキストを保存する工程と、

認識時に、各候補綴りパターンをマッチングして、任意の保存された多音節シーケンスに一致するか分かるようにする工程と、

一致した場合には、与えられた候補綴りパターンをキーとして、多音節シーケンスの前記拡張コンテキストおよび、対象綴りパターンを保存する工程と

を更に含む、請求項 1 に記載のデータ処理方法。

【請求項 1 4】

音節分けが単語位置表記を含み、オンセット最大化の原則を用いて実行される場合には、その音節分け標準発音パターンにアラインメントされた綴りパターンを含む発音辞書を保存する工程を、

更に含む、請求項 7 に記載のデータ処理方法。

【請求項 1 5】

生成記号ストリームは、コンピュータシステムのデジタル方式でプログラムされたロジックを用いて、テレビ番組のオーディオ入力から生成されるものであり、

前記生成記号ストリームの綴り表記を、前記テレビ番組を表示している表示装置に、該テレビ番組の字幕として表示させる工程を、

更に含む、請求項 1 に記載のデータ処理方法。

【請求項 1 6】

システムにおいて、

1つ以上のプロセッサと、

命令を保存するメモリと

を含み、前記命令は、前記 1つ以上のプロセッサによって実行された場合に、

特定の言語で話された単語から生成された生成記号ストリームを、音響処理システムから受信する処理と、

複数の生成パターンを、前記生成記号ストリームから抽出する処理と、

生成パターンから標準パターンへの 1つ以上のマッピングについての条件付き確率を含むものである、保存された生成 - 標準マッピングデータを用いて、候補音節、および、各前記候補音節の確率を、前記複数の生成パターンから生成する処理と、

10

20

30

40

50

1つ以上の候補音節から標準パターンへのマッピングについての条件付き確率を含むものである、保存された音節 - 綴りパターンマッピングデータを用いて、候補綴りパターン、および、各前記候補綴りパターンの確率を、前記候補音節から生成する処理と、

前記生成記号ストリームの綴り表記を、少なくとも部分的には各前記候補綴りパターンの前記確率に基づいて生成する処理と  
を行わせるものであるシステム。

【請求項 17】

前記生成記号ストリームは、母音間子音（IVC）単位および母音近傍（VN）単位を含む音素配列単位へと分割されるものであり、その分割は、該生成記号ストリームの各記号について、

3 記号バッファをゼロに、IVC 累積バッファをゼロとなるように初期化して、生成記号を順に前記 3 記号バッファに追加する処理と、

記号を前記 3 記号バッファに追加した後に、該 3 記号バッファの中央の記号が母音であり、したがって、それらの 3 つの記号は、VN を含むかを決定して、前記 VN を保存する処理と、

追加された記号が子音の場合は、前記子音を、前記 IVC 累積バッファに加える処理と、次に追加された記号が子音ではない場合は、前記記号を前記 IVC 累積バッファに、IVC として保存する処理と、

前記生成記号ストリームが続く限りは処理を継続する処理と

を順に行うことによって行われるものである、請求項 16 に記載のシステム。

【請求項 18】

前記生成記号ストリームは、母音間子音（IVC）単位および母音近傍（VN）単位を含む音素配列単位に分割された生成音素配列ストリームであり、前記候補音節を生成する工程は、前記生成音素配列ストリームを、標準音素配列ストリームにマッピングする処理を含み、

その処理は、

処理待ち行列を、空となるように初期化する処理と、

生成 IVC と生成 VN の交互のストリームを前記生成音素配列ストリーム内で受信する処理と、

アラインメントされた生成 IVC および標準 IVC の予め計算保存されたマッピングデータを用いる処理であって、生成 IVC を 1 つ以上の標準 IVC に、保存され関連付けられた条件付き確率と共にマッピングするものである処理と、

アラインメントされた生成および標準 VN の保存されたマッピングデータを用いる処理であって、生成 VN を 1 つ以上の標準 VN に、保存され関連付けられた条件付き確率と共にマッピングするものである処理と

によって行われるものである、請求項 16 に記載のシステム。

【請求項 19】

前記命令は、前記 1 つ以上のプロセッサによって実行された場合に、前記マッピングされた標準音素配列ストリームを更に処理する処理を、  
行わせ、その処理は、

各前記標準 IVC が、直前の標準 VN パターンの組の少なくとも 1 つの要素から整合して連続したものであるかを決定し、そのような標準 VN パターンの少なくとも 1 つとも整合しない標準 IVC を除外する処理と、

各前記標準 VN が、直前の標準 IVC パターンの組の少なくとも 1 つの要素から整合して連続したものであるかを決定し、そのような標準 IVC パターンの少なくとも 1 つとも整合しない標準 VN を除外する処理と、

整合しない IVC および VN を除外した後に、条件付き確率を再び正規化する処理と  
によって行われるものである、請求項 18 に記載のシステム。

【請求項 20】

前記候補音節を生成する処理は、

10

20

30

40

50

整合した標準 I V C パターンを、音節のコーダ オンセットシーケンスへと分割する処理と、

分割された整合した標準 I V C パターンのコーダ部分を用いて、残りの I V C パターンおよび V N パターンから増やされたオンセットを通る整合した経路を完成する処理であり、各前記整合した経路は候補音節を形成するものである処理と、

保存された音節表にない候補音節を除外する処理と、

各前記候補音節の確率を計算する処理と、

前記候補音節を保存する処理と

を含むものである、請求項 16 に記載のシステム。

【請求項 21】

各前記候補音節について確率を生成する処理は、

特定の生成記号の組に対応する各候補音節について、

前記候補音節についての前記 V N の確率を、前記生成 標準マッピングデータ の保存計算された確率から特定する処理、

前記候補音節のオンセットの確率を、前記オンセットがそれに基づきうる各標準 I V C についての前記生成 標準マッピングデータ の保存計算された確率に基づいて特定する処理と、

前記音節のコーダの確率を、前記コーダがそれに基づきうる各標準 I V C についての前記生成から標準へのマッピングの保存計算された確率に基づいて特定する処理と、

前記 V N の前記確率、前記オンセットの前記確率、および、前記コーダの前記確率を集計して、音節確率値とする処理と、

前記特定の生成記号の組について、前記候補音節の前記音節確率値の合計を正規化する処理と

を含むものである、請求項 20 に記載のシステム。

【請求項 22】

候補音節を綴りパターンにマッピングする処理は、

アラインメントされた標準音節分けした発音および綴り音節パターンを含むデータを保存する処理と、

待ち行列を、空となるように初期化して、前記候補音節を順に提供し、学習保存された音節分けした綴りパターンの組にマッピングする処理と、

前記パターンと関連付けられた単語位置コードを用いて、無効パターンを除外する処理と、

前記マッピングされた綴りパターンを保存する処理と

を含むものである、請求項 16 に記載のシステム。

【請求項 23】

前記命令は、前記 1 つ以上のプロセッサによって実行された場合に、音節位置制約、および、統計モデルを適用して、出力綴りパターンを選択する処理を、更に行わせ、その処理は、

綴りパターンを、待ち行列を通して、重なる音節のシーケンスが順次処理される複数の音節サイズの単位のスライディングウインドウの形態で順に処理する処理と、

綴りパターンの組を受信した時に、各綴りパターンの組に対応する  $n$  グラムモデルに保存されたデータを参照する処理と、

$n$  グラムモデルの条件付き確率を合計して、経路確率を生成する処理と、

前記経路確率を保存する処理と

によって行われるものである、請求項 16 に記載のシステム。

【請求項 24】

前記命令は、前記 1 つ以上のプロセッサによって実行された場合に、

アラインメントされた生成 パターン および標準 パターン を含むデータを保存する処理と、

前記保存された マッピングデータ の生成 パターン および標準 パターン の各組合せについて、前記標準パターンの確率を、該保存された マッピングデータ の前記組合せの発生回数

10

20

30

40

50

に対応する値と、異なる生成パターンの前記標準パターンの発生回数に対応する値との関数として、計算する処理と、

前記生成パターンおよび標準パターンの組合せを、対応する前記生成 - 標準マッピングデータの前記組合せの計算した確率と共に保存する処理とを更に行わせるものである、請求項 1.6 に記載のシステム。

【請求項 25】

前記命令は、前記 1 つ以上のプロセッサによって実行された場合に、  
アラインメントされた音節分け標準発音および綴りパターンを含むデータを保存する処理と、

前記保存されたマッピングデータの音節分け標準発音および綴りパターンの各組合せについて、前記パターンの確率を、該保存されたマッピングデータの前記組合せの発生回数に対応する値と、該保存されたマッピングデータの前記音節分け標準発音の発生回数に対応する値との関数として、計算する処理と、

前記音節分け標準発音および綴りパターンの組合せを、対応する前記生成 - 標準マッピングデータの前記組合せの計算した確率と共に保存する処理とを更に行わせるものである、請求項 1.6 に記載のシステム。

【請求項 26】

前記命令は、前記 1 つ以上のプロセッサによって実行された場合に、  
文字起こしされた特定の言語で話されたか、または、書かれた音声を含む総語数コーパスを保存する処理と、

特定の言語の単語の音節分けした発音および対応する綴りパターン含む発音辞書を保存する処理と、

前記発音辞書を用いて、前記総語数コーパスを、音節分けした綴りのチャンクコーパスに変換する処理と、

前記音節分けした綴りのチャンクコーパスを用いて、各 n グラムパターンについて関連回数を累積して、各候補綴り音節パターンについて確率を生成する処理とを更に行わせるものである、請求項 1.6 に記載のシステム。

【請求項 27】

前記命令は、前記 1 つ以上のプロセッサによって実行された場合に、  
文字起こしされた特定の言語で話されたか、または、書かれた音声を含む総語数コーパスを保存する処理と、

特定の言語の単語の音節分けした発音および対応する綴りパターン含む発音辞書を保存する処理と、

前記発音辞書を用いて、前記総語数コーパスを、発音音節コーパスに変換する処理と、

前記発音音節コーパスを用いて、各 n グラムパターンについて関連回数を累積して、各候補発音音節パターンについて確率を生成する処理と

を更に行わせるものである、請求項 1.6 に記載のシステム。

【請求項 28】

前記命令は、前記 1 つ以上のプロセッサによって実行された場合に、

訓練中に、与えられた候補綴りパターンが、対象綴りパターンに一致しない最も尤度が高い綴りパターンであり、

与えられた候補綴り音節パターンを含む多音節シーケンスの拡張コンテキストを保存する処理と、

認識時に、各候補綴り音節パターンをマッチングして、任意の保存された多音節シーケンスに一致するか分かるようにする処理と、

一致した場合には、与えられた候補綴りパターンをキーとして、多音節シーケンスの前記拡張コンテキストおよび、対象綴りパターンを保存する処理と

を更に行わせるものである、請求項 1.6 に記載のシステム。

【請求項 29】

前記命令は、前記 1 つ以上のプロセッサによって実行された場合に、

10

20

30

40

50

音節分けが単語位置表記を含み、オンセット最大化の原則を用いて実行される場合には、その音節分け標準発音パターンにアラインメントされた綴りパターンを含む発音辞書を保存する処理を、  
更に行わせるものである、請求項 2 1 に記載のシステム。

【請求項 3 0】

生成記号ストリームは、コンピュータシステムのデジタル方式でプログラムされたロジックを用いて、テレビ番組のオーディオ入力から生成されるものであり、

前記命令は、前記 1 つ以上のプロセッサによって実行された場合に、

前記生成記号ストリームの綴り表記を、前記テレビ番組を表示している表示装置に、該テレビ番組の字幕として表示させる処理を、

10

更に行わせるものである、請求項 1 6 に記載のシステム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本開示の一技術分野は、話者を問わないコンピュータによる自動音声認識である。他の技術分野は、記録されたオーディオ信号からの音声を表す記号を、視聴覚番組のグラフィカルテキスト字幕などで連続表示しうるテキストに自動変換する技術である。本開示は、概して、話された単語をテキストまたは他の記号表記に変換する処理に関し、音声の自動文字起こしを含む。

【背景技術】

20

【0 0 0 2】

本章で記載するアプローチは、追及しうるものであるが、必ずしも、既に考えられたり、追及されたりしたものとは限らない。したがって、別段の記載がない限りは、本章のいずれのアプローチも、単に、本章に含められたというだけで、従来技術であると見なされるべきではない。

【0 0 0 3】

ライブテレビ放送中に話された音声に対応するテキスト字幕の自動生成表示など、自動音声認識 (ASR) のいくつかの利用例において、単語の正確さが高いこと、単語の発話から、対応する文字起こしテキストが入手可能になるまでの遅延が小さいか、無いこと、更に、話された単語が予め定義された語彙リストにない場合でも、良好な結果を生成できることが必要である。更に、計算効率が高く、省エネルギーで、クラウドに接続しなくても、分散した計算装置上で完全な性能で機能しうる ASR システムも必要である。

30

【発明の概要】

【0 0 0 4】

本開示の目的は、話者が発声し、音響特性分析により記号形式に自動で割り当てられた音を表す記号の入力ストリームを、対応する文字起こしされた綴りパターンの出力ストリームに変換することである。発声された音は標準発音から大きく異なることが多いが、そのような場合でも、意図された単語に正確に文字起こしすることも目的とする。更に、システムの訓練中に用いた参照辞書になかった単語についても、妥当な文字起こしを行うこと、つまり、「語彙外」問題を軽減することも目的とする。更に、音声生成と、それに対応する文字起こしされた綴りパターン生成の間の遅延(「待ち時間」)を最小にすることも目的とする。更に、上記目的を、低い計算複雑性で実現することを目的とする。音節は、音声の基本時間構造単位である。本開示のいくつかの態様は、音節の周囲、および、その内部構造での処理工程を、分割、アラインメント、同期、マッピング、更に、モデル化を行うように構成して、上記目的を実現する。

40

【図面の簡単な説明】

【0 0 0 5】

【図 1】オンセット、核、および、コーダを含む音節構造を示している。

【図 2】IPA 記号と K l a t t e s e の対応を示す図である。

【図 3】多音節語の構造を示す。

50



【図 4】様々なストリームの関係を示す。

【図 5 A】例示的な訓練参考材料を示す。

【図 5 B】例示的な訓練参考材料を示す。

【図 6】音素配列パターンを発音辞書から学習する例示的な方法を示すフローチャートである。

【図 7】生成パターンから標準パターンへのマッピングを学習する例示的な方法を示すフローチャートである。

【図 8】候補音節から綴りパターンへのマッピングを学習する例示的な方法を示すフローチャートである。

【図 9】I V C パターンおよび V N パターンを、生成ストリームから抽出する例示的な方法を示すフローチャートである。

10

【図 10】候補音節の組を、標準パターンの組から生成する例示的な方法を示すフローチャートである。

【図 11】整合性制約の適用を示す図である。

【図 12】生成パターンの標準 I V C に対する確率の例を示し、整合性制約を用いた削減を含む。

【図 13】候補音節を綴りパターンにマッピングする例示的な方法を示すフローチャートである。

【図 14】位置制約および統計モデルを適用して、出力綴りパターンを選択する例示的な方法を示すフローチャートである。

20

【図 15】綴り候補のコンフュージョンネットワークを示す。

【図 16】音節構造を用いて自動音声認識を行う例示的な方法を示す。

【図 17】実施形態を行いうるコンピュータシステムを示すブロック図である。

【発明を実施するための形態】

【0006】

以下の記載において、本開示の完全な理解のために多数の具体的な詳細事項を示して説明する。しかしながら、本発明の実施形態は、これらの具体的な詳細事項を含まずに実施しうることが明らかだろう。他の場合において、本開示を不必要に曖昧にしないために、既知の構造および装置をブロック図で示している。実施形態を以下のような各章で開示する。

30

【0007】

1. 概観

2. 例示的な言語処理システムの概観

3. 音節構造の概観

4. 動作段階

4.1 訓練段階

4.1.1 発音辞書から音素配列パターンの学習

4.1.2 生成パターンから標準パターンへのマッピングの学習

4.1.3 標準音節と綴りパターンの関係の学習

4.1.4 音節発音空間で n グラムモデルの学習

4.1.5 音節綴り空間で n グラムモデルの学習

4.1.6 閾値およびオーバーライドの学習

4.2 認識段階

5. 使用例

6. 実施形態の利点

7. ハードウェア概略

40

【0008】

1. 概観

実施形態において、方法は、音素配列規則および制約を反映した音節および音声の他の音節サイズの単位を用いて、話された音の入力ストリームを表す記号および他の非音声記

50

号（生成ストリーム）を、標準発音のストリーム（標準ストリーム）に変換し、次に、標準ストリームを、綴りテキストの出力（綴りストリーム）に変換する。

【0009】

方法は、訓練段階および認識段階を含む。訓練段階は、（特別に構成された発音辞書、並びに、文字起こしされた単語および文のコーパス含む）訓練材料の組を用いて、（生成、標準、および、綴りの）3つのストリームから、音節およびそれらの構成要素、母音間子音（IVC）音素シーケンス、母音近傍（VN）音素シーケンス、および、音節サイズの綴り単位を含む言語音素配列単位またはパターンのうち、いくつかの疑似閉集合の組を、抽出する工程を含む。これらの単位またはパターンの各組は、音素、音節および単語の配列および分割を支配する規則によって制約される。次に、処理は、生成ストリームから抽出された各単位を、全ての容認標準単位と関連付けるか、マッピングさせて、各生成標準関連付けの条件付き確率を決定する。訓練材料を再び用いて、処理は、標準単位またはパターンを、容認綴りパターンと関連付けるか、マッピングさせて、条件付き確率を、各標準綴り組合せに指定する。更に、訓練材料を用いて、音節の発音および綴りの両方について、 $n$  グラムモデルを生成する。最後に、訓練工程は、学習し、容認されるが非常に低い確率の音節を破棄して、拡張コンテキストツールを生成して、前の工程で満足な結果が得られなかった場合を修正する。

10

【0010】

認識段階は、言語学上の単位の組、および、訓練段階で学習した3つのマッピングされたストリームの条件確率を用いる工程を含む。生成IVCおよび生成VNを、入力音響記号ストリームから抽出して、生成IVCおよびVNストリームまたは待ち行列を生成する。そのストリームは、学習した条件付き確率を用いて標準IVCおよびVNパターンの組にマッピングされて、標準IVCおよびVNストリームまたは待ち行列を生成する。標準IVCおよびVNのストリームから、学習した容認IVCおよびVNの組と整合しないIVCおよびVNパターンが除外され、条件付き確率が、再び正規化される。次に、標準ストリームは、候補音節のコーダ オンセットシーケンスの組に分割されて、そこから、候補音節のストリームが生成される。容認できない音節、つまり、音節表にない音節は除外されて、正規化された条件付き確率が、容認候補について、IVC、VN、コーダ、および、オンセットの学習した確率から生成される。次に、候補IVCパターンを、学習して関連付けられた綴りパターンの組にマッピングする。学習した音節位置制約と整合しないパターンは除外されて、その結果得られる容認綴りパターンのストリームに、スライディング $n$  グラムコンテキストモデルを適用して、経路確率を生成する。最も尤度が高い綴りパターンを、それが、学習した拡張コンテキストのオーバーライドの組の要素として特定されない限りは、出力し、特定された場合には、オーバーライドを出力する。

20

30

【0011】

2. 例示的な言語処理システムの概観

図5Aは、音声および非音声を表す記号ストリームを、綴りパターンの出力ストリームに変換するマッピングを訓練するための例示的なシステムを示している。

【0012】

計算システム500を、1つ以上のプロセッサコアを有するサーバクラスのコンピュータまたは他のコンピュータ、コプロセッサ、若しくは、他のコンピュータを用いて実現しうる。計算システム500は、物理的サーバコンピュータ、および/または、クラウドコンピューティングを介するものなど、データセンタに保存された仮想サーバインスタンスでありうる。更に、または、その代わりに、計算システム500は、スマートフォン、パーソナルコンピュータ、タブレットコンピューティング装置、PDA、ラップトップ、または、情報の送受信が可能で本明細書に記載の機能を行いうる任意の他の計算装置でありうる。

40

【0013】

図5Aは、単一の計算システム500を、分かり易い例を示すために個別の要素と示している。しかしながら、他の実施形態において、より多くの計算システムを用いて、本明

50

細書に記載の機能を行いうる。例えば、第1の計算システムは、訓練材料502を用いて、抽出、マッピング、および、音素配列単位の組の生成を行い、マッピング512を、第2の計算システムに送信しうる。第2の計算システムは、マッピング512を用いて、綴りパターンを、記号ストリームから生成しうる。更に、または、その代わりに、多数の計算システムを用いうる。

【0014】

計算システム500は、訓練材料502を含む。訓練材料502は、抽出およびマッピング512を生成するのに用いるデータを含む。図5Aの実施形態において、訓練材料502は、発音辞書504、平行に文字起こしされた音声音のコーパス506、および、総語数コーパス508を含む。他の実施形態において、多少の訓練材料を用いて、抽出およびマッピング512を生成しうる。

10

【0015】

発音辞書504は、特定の言語の単語の音節分けした標準発音および対応する綴りパターンを含む。本開示では、例として英語を用いているが、本明細書に記載の方法は、任意の言語で用いうる。

【0016】

平行に文字起こしされた音声音のコーパス506は、特定の言語の異なる音ストリームのアラインメントされた生成および標準発音を含む。

【0017】

総語数コーパス508は、特定の言語で話されたか、または、書かれた一連の句または文を文字起こししたものを含む。

20

【0018】

これらの3つツールの全ては、専有材料または商業的に入手可能なコーパス、若しくは、それらの組合せから構成されうる。

【0019】

訓練命令510は、計算システムのRAMなどのメインメモリの1つ以上のページの組を含みうるもので、そこに、実行可能な命令がロードされており、実行された時には、計算システムに、それらのモジュールについて本明細書に記載の機能および動作を行わせうる。例えば、訓練命令は、RAMのページの組を含み、それは、実行された場合には、生成発音パターンの各インスタンスの特定、および、そのパターンの1つ以上の標準発音パターンへのマッピングを行わせる命令を含みうる。

30

【0020】

命令は、CPU、GPU、Phiまたは他の処理チップの命令の組で機械実行可能なコードであり、JAVA（登録商標）、C、C++、OBJECTIVE-C、または、任意の他の人間が読むことができるプログラミング言語または環境で記述されたソースコードを元にコンパイルされたもののみか、若しくは、JAVASCRIPT（登録商標）、他のスクリプト言語、および、他のプログラミングソーステキストのスクリプトとの組合せでありうる。「ページ」という用語は、広義に、メインメモリ内の任意の領域のことを称することを意図し、システムで用いる具体的な用語は、メモリアーキテクチャまたはプロセッサアーキテクチャに応じて異なりうる。他の実施形態において、訓練命令510は、図5Aのシステムまたは別のリポジトリシステムで、不揮発性RAMまたはディスク記憶装置などの大容量記憶装置にデジタル方式で保存されたソースコードの1つ以上のファイルまたはプロジェクトも表しうるもので、それは、コンパイルまたはインタープリットされた場合に、実行可能な命令を生成し、命令が実行された場合に、コンピュータが、それらの命令について、本明細書に記載の機能または動作を行いうる。つまり、図面は、プログラマーまたはソフトウェア開発者が、後でコンパイルして実行可能にするか、または、コンピュータが実行するためにバイトコードまたは等価物にインタープリットするソースコードを構成配列する様子を示しうる。

40

【0021】

実施形態において、計算システム500は、訓練命令510を用いて、様々な言語処理

50

値 5 1 2 を生成する。これらは、発音辞書から学習した音素配列データの組を含み、それは、音節（音節表）、音節オンセット、音節コーダ、条件付き確率と関連付けられた I V C パターンおよび V N パターンの包括的な組を 1 つ以上含む。これらは、4 . 1 . 2 章に記載するような生成 I V C および V N パターンの標準 I V C および V N パターン 5 1 6 へのマッピング、並びに、標準音節の綴りへのマッピング 5 1 8 も含む。これらは、音節発音の n グラムモデル 5 2 0、および、音節綴りの n グラムモデル 5 2 2 を学習する工程も含む。これらは、4 . 1 . 6 章に記載するような音節確率閾値およびコンテキストのオーバーライド 5 2 4 を学習する工程も含む。

#### 【 0 0 2 2 】

計算システム 5 0 0 は、更に、マッピングを用いて、入力生成パターンから綴りパターンを生成するようにプログラムまたは構成されうる。更に、または、その代わりに、計算システム 5 0 0 は、マッピングを、ネットワークを介して異なる計算装置に送信するようにプログラムまたは構成されうる。ネットワークは、広義で、1 つ以上のデータ通信ネットワークの任意の組合せを表し、それは、ローカルエリアネットワーク、ワイドエリアネットワーク、インターネットワークまたはインターネットを含み、地上または衛星リンクを含む任意の有線または無線リンクを用いるものである。ネットワークは、計算システム 5 0 0 と第 2 の計算システムの間でのデータ交換を提供する任意の媒体または機構によって実現されうる。計算システム 5 0 0 と第 2 の計算システムは、直接（有線または無線）通信リンクも有しうる。計算システム 5 0 0 は、T C P / I P、B l u e t o o t h（登録商標）などの標準ネットワーク通信プロトコル、および、H T T P、T L S など、より高いレイヤのプロトコルを用いるようにプログラムまたは構成されうる。

#### 【 0 0 2 3 】

##### 3 . 音節構造の概観

音声は、調音器官（つまり、あご、舌、唇など）の構成を変化させて生成される。子音は、声道に沿ったどこかで、狭窄を形成することによって生成され、一方、母音は、もっと開放した構成で生成される。狭窄と開放を交互に行うことで、音声信号に変調パターンを生じて、それは、何か他の音響および音素配列の合図も用いて、音節に感覚的に分割される。音節は、音声の基本的時間構造単位である。訓練がほとんどなくても、子供は、「意味を成さない」単語（例えば、「n a p k a p i t y」）を含む未知の単語についてさえ、発話された音節数を数えうる。音節の「ビート」は、音声に特徴的なリズムを与えて、混同しうる代替りのものに焦点を当てた感覚処理が、各ビート内で、できるようにする。音節の内部構造（つまり、音素配列）が、その言語で容認される音素シーケンスを局所的に制約するので、認知効率は更に高められる。

#### 【 0 0 2 4 】

音素配列制約は、音節間の潜在的な分割点も限定する。単語の境は、音節の境の部分集合なので、音節分割の制約は、単語分割も制約する。音節は、多数の他の方法で、音声認知に貢献する。例えば、音節は、語の強勢の担体であり、更に、疑問文を平叙文から識別、句の境検出などに用いられるピッチパターンおよび時間延長という合図も担持する。

#### 【 0 0 2 5 】

本開示のいくつかの態様は、音節の構造を活用する。図 1 は、単音節語「s t r e n g t h s」の構造を示している。図 1 に示すように、音節は、3 つの部分である「オンセット」1 0 3、「核」1 0 4、および、「コーダ」1 0 5 を含む。オンセット 1 0 3 は、核の前のゼロ以上の子音のシーケンスである。核 1 0 4 は、単一の母音的な音である。コーダ 1 0 5 は、核の後のゼロ以上の子音のシーケンスである。図 1 において、更に、本開示を通して、発音 1 0 2 は、「K l a t t e s e」で表している。

#### 【 0 0 2 6 】

I P A 記号から K l a t t e s e への対応を、図 2 に示す。この表記を、利便性が高く、計算が容易なことから用いる。音素配列規則が支配する音節を、調音器官の特徴パターンまたは聴覚特徴パターンなどの任意の他の発音記号表記で画定しうる場合には、それらの表記を用いる。図 2 から分かるように、音素記号を、子音 2 0 0 または核 2 0 1 とし

10

20

30

40

50

て分類しうる。

【0027】

各言語は、オンセットで、コーダで、または、音節全体で生じうる記号シーケンスを制限する音素配列規則の組を有する。英語では、例えば、/str/は、コーダではなく、オンセットとして容認される。同様に、/GkTs/は、オンセットではなくコーダとして容認される。多音節語、および、話された単語のシーケンスにおいて、1つの音節のコーダの後に、次の音節のオンセットがくる。図3は、これを、2音節の単語「action」を用いて示している。ここで、コーダ オンセットシーケンス内の記号パターンを、IVC301パターンと称する。IVCパターン内の音節の境は、IVCパターンを分割した結果、有効コーダパターンおよび有効オンセットパターンが得られる場所だけで生じうる。

10

【0028】

図3を参照すると、潜在的には、IVCパターン301の/kS/は、/kS/、/kS/、または、/kS/に分割されうる。しかしながら、/kS/は、有効コーダでも、有効オンセットでもないため、唯一の有効な分割は、/kS/である。いくつかのIVCパターンについて、コーダ オンセットシーケンスを分ける1つより多くの有効な分け方がある。そのような場合に、辞書発音を音節分けする時に、実施形態は、有効オンセットである可能性があるものから最長のもので選択されるようにIVCを分割する「オンセット最大化の原則」(MOP)を用いる工程を含む。例えば、単語「modal」は、modal、または、modalに分割しうる。MOPは、前者の分け方を必要とする。「コーダ最大化の原則」(MCP)など、任意の他の整合した方法を用いる。いずれの方法を用いた場合にも、共通の組の様々な長さのIVCとなる。更に、オンセットとコーダの複雑さは、言語によって異なる。したがって、IVCの最大数、および、IVC内の子音の数も、言語、および、関連辞書が1つより多くの言語から単語を組み込む程度に応じて長さが異なる。

20

【0029】

VN302パターンは、音節の核の前の音素、核、および、核の後の音素を含む。オンセットおよび/またはコーダがnullの音節について、VNの最初および/または最後の記号を、音声の切れ目、または、隣接した音節からの音声と称しうる。

【0030】

音節構造を活用することで、いくつかの利点を提供する。音節、および、その内部構造は、容認音パターンについての局所決定的制約を提供する(例えば、IVCパターンは、有効コーダ オンセットシーケンスを含まなくてはならない、発話の最初のIVCは、有効オンセットでなくてはならないなど)。これらの制約と整合しない仮のパターンは価値がないとして、更なる処理から除外されて、認識時間で要求される計算の数を大きく削減し、有効パターンの確率を高めうる。

30

【0031】

容認コーダおよびオンセットパターンの組は、可能な音節分割点を非常に制約する。これにより、競合する仮の音節が、一時的にアラインメントされた組を形成するのを可能にする。通常の話す速度は、毎秒8未満の音節を生成し、各音節「ビート」の到着と同期した処理工程を、遅い速度においても適用するのを可能にする。十分に大きく適切に用意された発音辞書から学習した場合には、オンセット、コーダ、音節、IVCパターン、および、VNパターンは、擬似閉集合を形成する。言語の音素配列規則が変化することは、稀である。したがって、語彙外という問題は、音節に関する単位に基づく場合には、大きく削減される。多数の発音の変形は、単語の端近くのIVCパターン内で、または、音節の中心部分のVNパターンで生じる規則的変化である。規則的変化を学習することによって、候補標準形を、変更された生成形から回復しうる。

40

【0032】

各音節の発音は、1つ以上の綴りパターンに関連付けられている。図1に示すように、綴りパターン「strengths」100は、音節の発音/str'EGkTs/102

50

に関連付けられる。/ s t r ' E G k T s /などのいくつかの音節は、1つだけの関連付けられた綴りパターンを有する。他の音節は、多数の関連付けられた綴りパターンを有する。例えば、/ D ' E r /は、綴りパターン「t h e r e」、「t h e i r」、および、「t h e y ' r e」と関連付けられる。所定の音節について関連付けられた綴りパターンは、単語中の音節位置によって影響を受けることが多い。例えば、/ D ' E r /の「t h e y ' r e」としての綴りのマッピングは、/ D ' E r /が単音節語を表す場合にのみ生じ、一方、綴りのマッピング「t h e r e」は、単音節語について、または、多音節語、例えば、「t h e r e f o r e」の最初の音節として、正しいことがありうる。

#### 【0033】

音節位置の綴りのマッピングへの影響を活用するために、発音辞書は、音節位置によって決定されたコードを含む。観察された関連を、単音節語については、「0」と符号を付し、多音節語の最初の音節については、「1」と符号を付し、多音節語の次の音節については、「2」と符号を付している。図1において、「0」の位置コード101は、「s t r e n g t h s」の音節/ s t r ' E G k T s /との関連が、単音節語で生じることを示している。図3において、「1」の位置コード303は、綴りパターン「a c」の音節/ ' @ k /との関連が、多音節語の最初の音節で生じることを示し、「2」の位置コード304は、綴りパターン「t i o n」の音節/ S x n /との関連が、多音節語の最初の音節以外で生じることを示している。

#### 【0034】

音節毎の処理を用いることで、コンテキストモデリングおよび制約を、音響および綴りの両方の観点で同じ単位に適用して、曖昧さの排除を補完しうる。

#### 【0035】

##### 4. 動作段階

本開示の実施形態は、訓練と認識の2つの段階に分けられる。両方の段階は、段階的な生成、標準および綴りの3つの記号ストリームで行われる。図4は、ストリームの関係を示している。最初のストリームは、訓練段階と認識段階の両方で最初の入力である生成402ストリームである。生成ストリームは、話者によって生成されて、音響信号の自動処理によって記号形に変換された入力音シーケンスを記号で表したものである。入力ストリームは、生成I V C 4 0 3パターンおよびV N 4 0 1パターンに分割される。

#### 【0036】

訓練段階で学習したマッピングを、認識段階で用いて、各生成I V C 4 0 3およびV N 4 0 1パターンを対応する標準I V C 4 0 4およびV Nパターン400の組に、関連付けられた条件付き確率を用いて変換する。これらの標準I V CおよびV Nパターンは、次に、標準音節405にマッピングされ、それは次に、綴りのストリーム406にマッピングされる。但し、図4において、明瞭に示すために、マッピングされたパターンの組を、それらの組の単一の要素で表している。

#### 【0037】

##### 4.1 訓練段階

訓練段階において、システムは、オンセット、コーダ、音節、I V C、および、V Nの許容パターン、並びに、生成ストリームのそのような各パターンと標準空間の1つ以上の対応するパターンの組との関連付けを学習する。更に、関連付けられた綴りパターンの組を、各標準音節について学習し、条件付き確率を、そのような各関連付けられたパターンに指定する。音節単位に基づいたn グラムコンテキストモデルも、音節発音および音節綴りパターンの両方に基づいて学習する。許容パターンとn グラムモデルの様々な組を、訓練材料を参照して学習する。図5Aに示すように、訓練材料は、発音辞書504、パラレルに文字起こしされた音音声ストリームのコーパス506、および、大規模な総語数コーパス508を含む。図5Bに示すように、発音辞書504は、音節表532、V Nの組534、I V Cの組536、音節コーダの組538、および、音節オンセットの組540を含む音素配列単位530の組のソースである。発音辞書504は、総語数コーパス508にも適用されて、音節分けした綴りの単語コーパス526、および、発音音節コーパ

10

20

30

40

50

ス 5 2 8 を生成する。

【 0 0 3 8 】

発音辞書 5 0 4 を含む単語の主要部を、最初に、公的に入手可能か専有の 1 つ以上の単語コーパスから取得する。

【 0 0 3 9 】

これらの各単語の発音をアメリカ英語について用意する場合、発音は、その単語が独立に注意深く話された時の「標準米語」発音を表すことを意図する。標準発音に追加で、その代替りの発音も、その単語の一般的な発音について、辞書に含めうるもので、それは、音節数が変化したものか（例えば、「probably」について、/pr'abli/）、並びに/若しくは、更なるコードおよび/またはオンセットを導入したものでありうる（例えば、「admiral」について、/'@dmrɪl/は、英語の通常のオンセットではない/mr/を導入する）。頻繁に生じる単語を含めることに追加で、単音節語、他の方法では表されない有効音素配列パターンの例である単語、および、新しい音節綴りパターンを導入する単語の全てを含めるように特に努力する。

10

【 0 0 4 0 】

いくつかの単語は、1 つより多くの「正しい」スペリング（例えば、「ameba」または「amoeba」）を有するが、ほとんどの利用例について、標準スペリングが望ましい。したがって、辞書の各単語について、単一の綴りパターンを、「標準」として選択し、それ以外の全てを「代替りのスペリング」と見なす。固有名詞および固有形容詞など、大文字で始める必要のある単語について、大文字で始めることを標準スペリングに含む。

20

【 0 0 4 1 】

発音を、オンセット最大化の原則を用いて音節分けした後に、標準綴りを、綴りのシーケンスを発音した音節に指定することによって音節分けして、その分割が、発音した音節への最も密接な対応を保つようにする。これは、いくつかの場合において、一般的な辞書綴り音節分けと異なる。例えば、「passion」/p'@sɪn/は、辞書で典型的に見つけられる「pas sion」ではなく、「pa s sion」に音節分けされる。単語「mission」/m'ɪsɪn/も同様に、「mi s sion」ではなく、「mi s sion」に音節分けされる。この通常と異なる綴り音節分けは、音節/Sɪn/が、多数の単語の綴りパターン「s sion」と関連付けられることを明らかにする。

【 0 0 4 2 】

辞書への入力が高フン付けされた複合語の場合、ハイフン記号は、複合語の最初の部分の最後の音節の綴りの部分になりうる。例えば、複合語「Anglo American」において、音節/glo/は、位置コード「2」を有する綴りパターン「glo」と関連付けられる。ハイフン付けされた単語において、ハイフンの後の音節の位置コードは、ハイフンの後に 1 つだけの音節がある場合には、「0」と設定し、ハイフンの後に 1 より多くの音節がある場合には、「1」と設定する。

30

【 0 0 4 3 】

但し、辞書は、いくつかの外来語（例えば、「schmuck」）および非英語名を含みうることに留意すべきである。したがって、学習した音素配列規則は、「純粋な」英語について定義された規則に厳密に限定されるものではない。更に、システムを英語を用いて記載しているが、本発明のアプローチを、他の言語の処理にも、それらの言語の参照辞書を用いて利用しうる。

40

【 0 0 4 4 】

4 . 1 . 1 発音辞書から音素配列パターンの学習

オンセット、コード、音節（つまり、「音節表」）、IVCパターン、および、VNパターンの組を、発音辞書 5 0 4 を参照して学習する。辞書の各単語は標準発音を有し、代替りの発音も有しうる。図 6 は、音素配列パターンを発音辞書から学習する例示的な方法を示すフローチャートである。図 6 を参照すると、オンセット、コード、音節、IVCパターンおよびVNパターンの組は、nullに設定される 6 0 0。発音辞書の各発音について、最初の母音の前の子音シーケンスは、オンセットの組に追加され 6 0 2、最後の母

50

音の後の子音シーケンスは、コードの組に追加される。辞書に未処理の入力がない場合には、IVCパターンの組を、各コードを各オンセットとシーケンスにすることによって生成する604。

【0045】

いくつかのコードおよび/またはオンセットは、単語内部だけで生じることが可能なので、生成したIVCパターンを、発音辞書の単語内部のIVCパターンと比べてチェックする。その組にない任意の単語内部のIVCパターンが見つかった場合には、以前に知られていなかったコードおよび/またはオンセットを、各々の組に追加し、IVCパターンの組を再び生成する605。

【0046】

工程606において、オンセット最大化の原則(MOP)を用いて、各発音を音節に分けて、音節を音節表に追加する。工程607において、辞書の各単語の発音の各母音について、VNパターンを抽出する。VNパターンは、核の前の記号、核の記号、および、核の後の記号の3つの記号シーケンスから構成される。核が、発音の最初の記号の場合、切れ目記号「#」をVNの最初の位置に追加することによって、VNパターンを拡張する。

【0047】

更なるVNパターンは、核が最初の記号の場合に、VNの最初の位置に各単語の発音の最後の記号を反復的に追加することによって生成される。同様に、核が単語の発音の最後の記号の場合には、切れ目記号「#」をVNの最後の位置に追加することによって、VNを拡張し、更なるVNパターンは、VNの最後の位置に各単語の発音の最初の記号を反復的に追加することによって生成される。オンセットおよび/またはコードパターンが子音を含まない場合には、VNは多音節または多数の単語に及びるので、核が発音の最初または最後の位置の場合には、更なるVNパターンの生成が必要である。このように生成されたVNパターンは、全ての可能な単語のシーケンスを説明する。

【0048】

4.1.2 生成パターンから標準パターンへのマッピングの学習

話された単語の発音は標準から異なることが多い。多くの違いは、音節の端近くで(つまり、IVCパターン内で)生じる。例えば、「Did you see her?」という発話において、最初の2つの単語は、標準の/dɪdyu/の代わりに/dɪju/と話されうる。/dy/の代わりに/j/と話されるのは、調音器官構成が関わって、規則的に生じるものである。他の違いは、コンテキストの音節中心部分への影響により生じる。

【0049】

本開示の一態様は、各生成IVCまたはVNパターンと、それと対応する尤もらしい標準IVCまたはVN対象パターンとのマッピングを学習することである。それは、アラインメントされた生成および標準音音声ストリーム506を含む訓練材料を参照することによって行われる。対応する標準IVCまたはVNパターンを有する各生成IVCまたはVNパターンの各観察回数を数える。関連回数を数えることで、観察したパターンを考慮した各標準パターンの条件付き確率が推定される。

【0050】

図7は、生成パターンから標準パターンへのマッピングを学習する例示的な方法を示すフローチャートである。更に、図7は、生成パターンから標準パターンへの条件付き確率マッピングを学習する処理を示している。各IVCまたはVNパターンを、そのようなパターンの例が訓練材料になくても、標準的に生成しうるということが知られている。したがって、工程700において、生成パターンの組を初期化して、標準パターンの組の各要素を含むようにする。対応する標準パターンと対にされた各生成パターンの関連回数は、1となるように初期化される。

【0051】

次の工程701において、生成ストリームで残りのIVCまたはVNパターンを特定する間に、標準ストリームで対応するIVCまたはVNパターンを特定する。工程702において、各観察した生成パターンと標準パターンの対について、関連回数を1加える。工

10

20

30

40

50



程 7 0 3 において、各対について平滑化された条件付き確率推定値を、以下のように計算する：

A は、生成パターンの観察数であり、

B は、対象パターンが生成パターンと関連付けられた回数であり、

C は、生成パターンと関連付けられた異なる対象パターンの数である場合、

【 0 0 5 2 】

【数 1】

$$\text{確率(対象パターン|生成パターン)} = \frac{B}{(A+C)} \quad \text{式. 1}$$

10

【 0 0 5 3 】

以下のような他の平滑化技術も用いる：

【 0 0 5 4 】

【数 2】

$$\text{確率(対象パターン|生成パターン)} = \frac{B+1}{A+2} \quad \text{式. 2}$$

【 0 0 5 5 】

4 . 1 . 3 音節と綴りパターンの関係の学習

20

既に記載したように、発音辞書 5 0 4 は、音節分けした発音および対応する綴りパターンを含む。図 8 は、候補音節の綴りパターンへのマッピングを学習する例示的な方法を示すフローチャートである。更に、図 8 は、単語中の音節位置を考慮した各綴りパターンの条件付き確率を学習するのに用いる処理を示している。マッピングを null となるように初期化した後に 8 0 0、残りの未処理入力がある間は 8 0 1、発音辞書における各発音の各音節 綴りの対を検査して、既にマッピングリストに加えられているか分かるようにする 8 0 2。加えられている場合には、関連回数を 1 加えて、工程 8 0 1 で処理を継続する。加えられていない場合には、新しい対を、関連回数 1 で、追加し 8 0 3、更に、工程 8 0 1 で処理を継続する。工程 8 0 4 において、処理すべき入力が残っていない場合には、関連回数を用いて、条件付き確率を以下のように計算する：

30

A は、音節発音の観察数であり、

B は、綴りパターンが音節発音と関連付けられた回数であり、

C は、音節発音と関連付けられた異なる綴りパターンの数である場合、

【 0 0 5 6 】

【数 3】

$$\text{確率(綴りパターン|生成パターン)} = \frac{B}{(A+C)} \quad \text{式. 3}$$

【 0 0 5 7 】

以下のような他の平滑化技術も用いる：

【 0 0 5 8 】

【数 4】

$$\text{確率(綴りパターン|生成パターン)} = \frac{B+1}{A+2} \quad \text{式. 4}$$

40

【 0 0 5 9 】

4 . 1 . 4 音節発音空間で n グラムモデルの学習

近傍の音節の発音のコンテキストの影響を、音節発音の n グラムモデルを訓練するこ

50

とによって取得する。実施形態において、3 グラムモデルを訓練する。これらのモデルを、大規模な総語数コーパス、文字起こしされた音声または書かれたテキストから取得した一連の文または句を用いて訓練する。総語数コーパスは、発音辞書を用いて、発音音節コーパスに変換されて、標準および任意の代わりに発音の両方を含む各単語の音節分けした発音を検索するようにする。n個の音節からなる各シーケンスは、n グラムを表し、本実施形態において、3つの音節からなる各シーケンスは、3 グラムを表す。次に、各3 グラムパターンについて、関連回数を累積する。回数を累積した後に、回数は、生成パターンから標準パターンへの条件付き確率を計算するのに用いた技術と同様の平滑化技術を用いて、確率に変換される。

【0060】

#### 4.1.5 音節綴り空間でn グラムモデルの学習

近傍の音節の綴りのコンテキストの影響も、n グラムモデルを訓練することによって取得する。実施形態において、3 グラムモデルを訓練する。これらのモデルを、大規模な総語数コーパスを用いて訓練する。総語数コーパスを、発音辞書を用いて、音節分けした綴りのチャンクコーパスに変換し、各単語の音節分けした綴りを検索するようにする。次に、各n グラムパターンについて、関連回数を累積する。回数を累積した後に、回数は、生成パターンから標準パターンへの条件付き確率を計算するのに用いた技術と同様の平滑化技術を用いて、確率に変換される。

【0061】

#### 4.1.6 閾値およびオーバーライドの学習

候補音節を構成して確率を指定する時に、いくつかは非常に低い確率を有する。実際、非常に低いので、綴りパターンにマッピングされた後に、これらの音節は選択されることがない。訓練後に、システムを試験して、その値より低い場合には特定の音節が出力結果に関わることがない確率閾値を決定する。この閾値を記録し、認識時に用いて、出力に影響しない候補を除外するようにする。

【0062】

曖昧さのほとんどは、局所的に解決されうる。曖昧さには、コンテキストを拡張することによって解決しうるものがある。曖昧さには、コンテキストを拡張しても解決されないものもある。むやみに大きいn グラムを用いて結果を改良する代わりに、短いものでは失敗する場合だけ、より大きいn グラムを生成する。

【0063】

学習中に、「最も尤度が高い」綴りパターンが対象パターンに一致しない場合には、システムは、「オーバーライド」を生成する。これは、「誤り」を「キー」として記録し、次に、その誤りを修正するであろう拡張コンテキスト(特定の最大数の音節または単語まで)を特定する。そのような拡張コンテキストを見つけた場合は、誤りキー、関連付けられたコンテキスト、および、正しい綴りパターンを記録する。この手法を、システム展開後にも採用して、システムの出力を改良するフィードバックを可能にしうる。

【0064】

#### 4.2 認識段階

図9は、IVCパターンおよびVNパターンを生成ストリームから抽出する例示的な方法を示すフローチャートである。認識中に、図9に示した処理を用いて、入力生成記号ストリーム402を処理して、生成IVC403およびVN401パターンを抽出し、それらを、各々、標準IVC404およびVN400にマッピングする。抽出処理は、丁度受信した記号(B3)、その直前に受信した記号(B2)、および、その直前に受信した記号(B1)を含む3 記号バッファを維持する。IVC累積バッファも維持される。

【0065】

任意の生成記号を受信する前に、3 記号バッファは、空となるように初期化される900。次に、IVC累積部をnullに設定する901。次の記号を受信し902、それは、3 記号バッファでB3になる903。次に、3 記号バッファを検査して、VNパターンを含むか分かるようにする904。含む場合には、VNパターンを待ち行列1にプ

10

20

30

40

50

ッシュする 905。受信した記号 B3 を検査して、子音かどうか分かるようにする 906。子音の場合には、IVC 累積部に加えられて 907、次の生成記号を受信しながら、処理が継続される 902。そうでない場合は、累積した IVC (null でありうる) を待ち行列 1 にプッシュして 908、次の生成記号を受信しながら、処理が継続される 902。  
【0066】

図 10 は、候補音節の組を標準パターンの組から生成する例示的な方法を示すフローチャートである。待ち行列 1 に IVC および VN パターンが到着すると、図 10 に示すように処理される。処理開始前に、待ち行列は空となるように初期化される 1001。次のパターンを受信し 1002、それが IVC パターンであるかどうかに基づいて処理される 1003。生成 VN パターンの場合には、標準 VN パターンの組にマッピングされる 1004。有効であるには、標準 VN パターンは、前の標準 IVC パターンの組のいくつかの要素から整合して連続したものを形成しなくてはならない 1005。整合しない VN パターンは、更なる処理から除外される。丁度受信したパターンが生成 IVC パターンの場合には、標準 IVC パターンの組にマッピングされる 1006。有効であるには、標準 IVC パターンは、前の標準 VN パターンの組のいくつかの要素から整合して連続したものを形成しなくてはならない。整合しない IVC パターンは、更なる処理から除外される 1007。残りの IVC パターンの要素への整合した連続を有さない、いずれの VN パターンも除外される。

【0067】

図 11 は、経路整合性制約の適用を詳細に示している。1100、1101、1102 は、待ち行列 1 に VN および IVC パターンが交互に到着するのを示している。VN および IVC パターンの有効候補マッピングは、到着パターンを通して整合経路を形成する。有効であるには、到着する IVC パターン 1101 は、その IVC パターンの直前に到着した VM パターン 1100 の残りの組の 1 つ以上の要素と整合しなくてはならない。VN パターン 1100 と IVC パターン 1101 の間の連続経路の整合性は、パターンの「重なった」部分 1106 を検査することによって決定される。具体的には、VN パターン 1100 の最終記号は、切れ目記号、母音記号または子音記号である。最終記号が切れ目記号または母音の場合には、有効な IVC の連続は、null コーダパターンを有するものへと制約される。子音の場合には、有効な IVC パターンの連続は、それと同じ子音で始まるものへと制約される。

【0068】

それについての有効な連続がない VN パターンは、除外される。同様に、少なくとも 1 つの残りの VN パターンからの有効な連続ではない IVC パターンも除外される。VN パターンが待ち行列 1 に到着する時、保持されるには、VN パターンは、前の候補 IVC パターンの組の残りの要素の 1 つ以上からの有効な連続を形成しなくてはならない。上記処理と同様に、整合性制約を、IVC および VN パターン 1107 の端の記号を検査することによって適用し、但し、この時には、マッチングは、IVC の最後の記号および VN の最初の記号が関わるものである。VN の最初の記号が子音の場合には、有効であるには、少なくとも 1 つの IVC の最後の記号と一致しなくてはならない。そうでない場合には、null オンセットパターンを有する IVC の連続にすぎない。

【0069】

再び、図 10 を参照すると、整合 IVC パターンがコーダ オンセットシーケンスに分割されると 1008、コーダ部分を用いて、残りの前の IVC パターンおよび残りの前の VN パターンから分割されたオンセットを通る整合経路を完成させる。各整合経路は、候補音節を形成する 1009。候補音節が音節表にない場合には、その候補音節は除外される 1010。容認候補音節を、待ち行列 2 にプッシュする 1011。パターンマッピング 1004、1006 により、生成パターンを考慮した各標準パターンマッピングの条件付き確率が推定される。IVC パターンがコーダ オンセットシーケンスに分割されると、IVC パターンのこれらの構成要素は、IVC パターンの確率を引き継ぐ。多数の残りの IVC パターンの分割が、同じコーダまたは同じオンセットパターンを生じる場合、共通

10

20

30

40

50

するパターンの確率は、それらを含む IVC パターンの確率の合計として推定される。オンセット、コーダ、および、VN の組の確率は、各々、合計で 1 に正規化される。音節候補 405 を形成する時に、それらのオンセットパターン、VN パターン、および、それらのコーダパターンの確率の積として計算した確率を指定する。次に、音節候補の組の確率は、合計で 1 に正規化される。次に、音節候補に指定された確率は、 $n$  グラムコンテキストモデルを適用することによって更に調節される。

#### 【0070】

図 12 は、整合性制約を適用することで、IVC の認識がいかに簡単になるかを例示的に示している。この例は、単一音素生成 IVC / T / の例である。訓練処理は、広範囲の条件付き確率 1201 を有する生成 IVC / T / 1200 と関連付けられた 58 の異なる標準 IVC を生成した。整合性制約の適用後に、可能性のある標準 IVC パターンの組は、かなり少なく 14 に削減され、つまり、約 76 % 削減される。次に、残りの IVC パターンの関連付けられた条件付き確率 1203 は、合計で 1 に正規化される。

10

#### 【0071】

次に、各候補音節 405 と関連付けられた綴りパターン 406 の組を生成する。図 13 は、候補音節を綴りパターンにマッピングする例示的な方法を示すフローチャートである。図 13 において、待ち行列 2 は空となるように初期化される 1300。候補音節の組を待ち行列 2 に受信し 1301、綴りパターンの組にマッピングする 1302。パターンと関連付けられた位置コード 101、303、304 を用いて、無効パターンを除外する 1303。残りのマッピングされたパターンを、待ち行列 3 にプッシュする 1304。

20

#### 【0072】

図 14 は、音節位置制約および統計モデルを適用して、出力綴りパターンを選択する処理を示している。待ち行列 3 を用いて、コンテキストで音節と関連付けられた綴りのチャンクを処理する。その処理は、音節綴りパターンの「N」個の直近の候補の組の履歴を含むスライディングウィンドウとして考える。パターンの条件付き確率、および、それらのコンテキストの確率を用いて、最も尤度が高いパターンを特定して、スライディングウィンドウの「最も古い」端で出力する。

#### 【0073】

実施形態において、スライディングウィンドウのサイズは、7 音節に設定される。他の実施形態は、コンテキストと待ち時間の間で異なるトレードオフで、異なる長さを用いる。待ち行列 3 は、空となるように初期化される 1400。次の綴りパターンの組を受信する 1401。シフトさせた  $n$  グラムコンテキストモデルを用いて、経路確率を生成する。次に、シフトさせた  $n$  グラムモデルの合計を適用する 1402。丁度出力されようとするパターンが、オーバーライドの組のキーとして見つけられ、かつ、コンテキストがオーバーライドと整合する場合は、オーバーライドと関連付けられた出力パターンを出力し、そうでない場合には、「最も尤度が高い」パターンを出力する 1403。

30

#### 【0074】

図 15 は、例示的な綴り候補のコンフュージョンネットワークを示している。3 音節クロック期間後に、3 グラムモデル 1506 を、パターンに、音節クロック期間 1500、1501、1502 で適用する。綴り候補が音節クロック期間 1503 に到着すると、3 グラムモデル 1507 を、パターンに、音節クロック期間 1501、1502、1503 で適用する。綴りパターンの各新しい組が到着する時に 1504、1505、3 グラムモデル 1508、1509 を各々適用する。

40

#### 【0075】

いくつかの状況において、生成ストリームを生成する音響認識部は、具体的な音素分類を高い信頼性で特定できないことがある。そのような場合、ストリームの記号は広い音声分類を示しうる。これが生じた場合には、標準パターンの組は、広い分類の要素である各具体的分類を表す記号を置き換えることによって、マッピングを繰り返すことによって生成された組の組合せとなる。例えば、音響プロセッサが、音素セグメントが / m / か / n / かを高い信頼性で決定できなかった場合には、広い分類 { / m / , / n / } を表す

50

記号を生成ストリームに挿入することになるだろう。生成から標準へのマッピングは、次に、別々のマッピングの組合せとして、マッピングされた標準 IVC パターンの組について、広い分類の記号の代わりに、/m/ および /n/ を IVC に挿入して置き換える。

【0076】

#### 5. 使用例

図16は、音節構造を用いて自動音声認識を行う例示的な方法を示している。

【0077】

工程1602において、オーディオ入力から生成された生成記号ストリームを受信する。例として、計算システムは、テレビ番組などを介して、オーディオ入力を受信する。音響認識部を用いて、生成記号ストリームを、オーディオ入力から生成しうる。計算システムは、処理のための生成記号ストリームを受信しうる。例えば、計算システムは、テレビ番組の字幕を生成するようにプログラムまたは構成されうる。

10

【0078】

工程1604において、4.1.1章および図6のように学習した生成 IVC および VN を、生成記号ストリーム内で、4.2章および図9で記載したように特定する。

【0079】

工程1606において、4.2章および図10で記載したように、生成 IVC および VN は、標準 IVC および VN にマッピングされ、4.1.2章および図7で記載した訓練処理で学習した音素配列パターンおよび条件付き確率を用いる。工程1608において、4.2章および図11で記載したように、経路整合性制約を用いて、整合しない IVC および VN を除外する。任意の単一の生成 IVC または VN について、複数の異なる標準 IVC および VN を生成しうる。

20

【0080】

工程1610において、4.2章および図10で記載したように、整合した IVC は、コーダ オンセットシーケンスに分割される。工程1612において、4.2章および図11で記載したように、コーダを用いて、候補音節を生成し、更に、音節表にない候補音節を除外する。

【0081】

工程1614において、4.2章に記載し、図12に示したように、各候補音節の確率を生成する。各候補音節の確率は、構成要素であるオンセット、コーダおよび IVC の学習した確率から導かれ、容認できない部分を除外した後に正規化し、更に、N グラムモデルを通して調節される。

30

【0082】

工程1616において、4.2章および図13に記載したように、候補音節は、4.1.3章および図8に記載したように訓練処理で学習した関連付けを通して、綴りパターンにマッピングされ、更に、工程1618において、整合しない綴りパターンは除外される。単一の候補音節を、多数の綴りパターンにマッピングしうる。

【0083】

工程1620において、4.2章および図14に記載したように、各綴りパターンの確率を、4.1.4章および4.1.5章に記載したような訓練処理で学習した構成要素である音節の確率を用いて生成する。

40

【0084】

工程1622において、4.2章および図14に記載したように、最も尤度が高い綴りパターンを、それが4.1.6章に記載したような訓練処理で学習した拡張コンテキストのオーバーライドの組のキーと一致しない限りは、出力し、一致する場合は、オーバーライドを出力する。他の実施形態において、n個の最も尤度が高い綴りシーケンスについて、辞書を再び参照し、完全に一致するものが特定された場合には、その一致したものを出力する。

【0085】

#### 6. 実施形態の利点

50

本明細書に記載のコンピュータによる処理は、マシンラーニング技術と言語学の知識を新しく、従来にはない組合せで用いて、生成された音声から抽出した音素および言語学記号を、標準綴りの正しいテキストに変換する。この自動音声テキスト変換システムの明確に異なる「バックエンド」の言語処理部は、現在の技術である「言語モデル」より計算効率が非常に高い。本発明は、効率的な「フロントエンド」の音響処理と組み合わせられた場合に、話された単語を、分散した装置で、高い正確さと低い待ち時間で連続して文字起こしすることが可能になりうる。それは、自動音声認識を採用する全ての目的で価値のあるツールとなるもので、限定するものではないが、ボイス検索、分散した装置および機器のボイスコマンド制御、および、ボイスディクテーションを含む。正確なリアルタイムの字幕およびサブタイトルを記録またはライブ映像に生成するのに、特に有用である。

10

【0086】

ライブ放送についての字幕は、高い正確さと、話された単語と表示された字幕の間でほとんど遅延がないか、全く遅延がないこと、更に、予め定義した語彙リストの部分にはないか、または、異なるアクセントで話された単語の字幕を生成できることが要求される。人間には、そのような字幕を、高い正確さ、または、高速で生成するのは困難である。本明細書に記載の規則を採用することによって、コンピュータは、受信した音ストリームを分割し、規則の特定の組を用いて最適な候補単語を特定し、更に、候補単語を、高い正確さで、高速に出力し、それにより、文字起こしされた出力結果を、ライブ放送でリアルタイムに再生するのを可能にする。更に、ニュース番組などのライブ番組では、固有名詞など、それより前には見たことのない単語が使われることが少なくない。現在の技術であるASRシステムは、その辞書に含まれた単語だけを出力できるものであり、それは、これらの見たことのない単語をいつも間違えることを意味する。本発明は、前に見たことのない単語を、音声から近似させて出力することが可能である。

20

【0087】

#### 7. ハードウェア概略

一実施形態によれば、本明細書に記載の技術は、1つ以上の専用計算装置によって実施される。専用計算装置は、ハードワイヤードされて技術を行うか、若しくは、技術を行うように持続的にプログラムされた1つ以上の特定用途向け集積回路(ASIC)またはフィールドプログラマブルゲートアレイ(FPGA)などのデジタル電子装置を含むか、若しくは、ファームウェア、メモリ、他の記憶装置、または、それらの組合せのプログラム命令に従って技術を行うようにプログラムされた1つ以上の汎用ハードウェアプロセッサを含みうる。更に、そのような専用計算装置は、カスタムハードワイヤードロジック、ASIC、または、FPGAを、カスタムプログラミングと組み合わせて、技術を実施しうる。専用計算装置は、デスクトップコンピュータシステム、ポータブルコンピュータシステム、手に持つ装置、ネットワーク装置、若しくは、ハードワイヤードおよび/またはプログラムロジックを組み込んで技術を行う任意の他の装置でありうる。

30

【0088】

例えば、図17は、実施形態を行いうるコンピュータシステム1700を示すブロック図である。コンピュータシステム1700は、バス1702、または、情報を通信する他の通信機構、並びに、バス1702と連結されて情報を処理するハードウェアプロセッサ1704を含む。ハードウェアプロセッサ1704は、例えば、汎用マイクロプロセッサでありうる。

40

【0089】

コンピュータシステム1700は、バス1702に連結されて、情報およびプロセッサ1704によって実行される命令を記憶するランダムアクセスメモリ(RAM)または他の動的記憶装置などのメインメモリ1706も含む。メインメモリ1706は、プロセッサ1704によって命令が実行される間に、一時変数または他の中間情報を記憶するのにも用いられうる。そのような命令は、プロセッサ1704にアクセス可能な非一時的な記憶媒体に記憶された場合、コンピュータシステム1700を、命令に指定された動作を行うようにカスタマイズされた専用マシンにする。

50

## 【 0 0 9 0 】

コンピュータシステム 1700 は、読出し専用メモリ (ROM) 1708、または、バス 1702 に連結されて静的情報およびプロセッサ 1704 のための命令を記憶する他の静的記憶装置を更に含む。磁気ディスク、光ディスク、または、固体ドライブなどの記憶装置 1710 が備えられ、それは、情報および命令を記憶するためにバス 1702 に連結される。

## 【 0 0 9 1 】

コンピュータシステム 1700 は、バス 1702 を介して、陰極線管 (CRT) など、コンピュータユーザに情報を表示する表示装置 1712 に連結されうる。英数字および他のキーを含む入力装置 1714 は、プロセッサ 1704 に情報およびコマンド選択を伝えるためにバス 1702 に結合されている。他の種類のユーザ入力装置は、マウス、トラックボール、または、カーソル方向キーなどのカーソル制御 1716 であり、方向情報およびコマンド選択をプロセッサ 1704 に伝え、表示装置 1712 上のカーソル移動を制御する。この入力装置は、典型的には、2 軸である第 1 の軸 (例えば、x) および第 2 の軸 (例えば、y) に 2 自由度を有し、平面内の位置を特定可能である。

10

## 【 0 0 9 2 】

コンピュータシステム 1700 は、本明細書に記載の技術を、カスタマイズされたハードワイヤードロジック、1 つ以上の ASIC または FPGA、ファームウェア、並びに / 若しくは、プログラムロジックを用いて実施しうるもので、それらは、コンピュータシステムとの組合せでコンピュータシステム 1700 を専用マシンにするか、または、プログラムする。一実施形態によれば、本明細書の技術は、コンピュータシステム 1700 によって、メインメモリ 1706 に含まれた 1 つ以上の命令の 1 つ以上のシーケンスを実行するプロセッサ 1704 に応じて行われる。そのような命令は、記憶装置 1710 などの他の記憶媒体からメインメモリ 1706 に読み込まれうる。メインメモリ 1706 に含まれる命令のシーケンスを実行することで、プロセッサ 1704 に、本明細書に記載の処理工程を行わせる。代替の実施形態において、ハードワイヤード回路を、ソフトウェア命令の代わりに、または、それとの組合せで用いうる。

20

## 【 0 0 9 3 】

「記憶媒体」という用語は、本明細書で用いるように、データ、および / または、機械を特別な態様で動作させる命令を記憶する任意の非一時的な媒体のことを称する。そのような記憶媒体は、不揮発性媒体、および / または、揮発性媒体を含みうる。不揮発性媒体は、例えば、光学ディスク、磁気ディスク、または、記憶装置 1710 などの固体ドライブを含む。揮発性媒体は、メインメモリ 1706 などの動的メモリを含む。記憶媒体の一般的な形態は、例えば、フロッピーディスク、フレキシブルディスク、ハードディスク、固体ドライブ、磁気テープ、または、任意の他の磁気データ記憶媒体、CD ROM、任意の他の光学データ記憶媒体、孔パターンを有する任意の物理的媒体、RAM、PROM、および、EPROM、FLASH EPROM、NVRAM、任意の他のメモリチップまたはカートリッジを含む。

30

## 【 0 0 9 4 】

記憶媒体は、送信媒体と別個のものであるが、共に用いうる。送信媒体は、記憶媒体間で情報の伝達に関わる。例えば、送信媒体は、同軸ケーブル、銅線、および、光ファイバを含み、バス 1702 を含むワイヤを。送信媒体は、電波および赤外線データ通信中に生成されるものなど、音響または光波の形態でもありうる。

40

## 【 0 0 9 5 】

様々な形態の媒体が、1 つ以上の命令の 1 つ以上のシーケンスをプロセッサ 1704 に実行のために搬送するのに関わりうる。例えば、命令は、最初に、磁気ディスク、または、遠隔コンピュータの固体ドライブに担持されうる。遠隔コンピュータは、命令を動的メモリにロードして、命令を、電話回線を通してモデムを用いて送信する。コンピュータシステム 1700 のローカルモデムは、電話回線でデータを受信し、赤外線送信器を用いて、データを赤外線信号に変換しうる。赤外線検出器は、赤外線信号に担持されたデータを

50

受信し、適切な回路は、データをバス 1702 に載せうる。バス 1702 は、データをメインメモリ 1706 に搬送し、プロセッサ 1704 は、そこから命令を読み出して実行する。メインメモリ 1706 によって受信された命令は、任意で、プロセッサ 1704 による実行の前または後に、記憶装置 1710 に記憶されうる。

【0096】

コンピュータシステム 1700 は、バス 1702 に連結された通信インターフェイス 1718 も含みうる。通信インターフェイス 1718 は、ローカルネットワーク 1722 に接続されたネットワークリンク 1720 に連結する双方向データ通信を提供する。例えば、通信インターフェイス 1718 は、統合サービスデジタル通信網 (ISDN) カード、ケーブルモデム、衛星モデム、または、対応する種類の電話回線へのデータ通信を提供するモデムでありうる。他の例として、通信インターフェイス 1718 は、互換性のあるローカルエリアネットワーク (LAN) へのデータ通信接続を提供する LAN カードでありうる。無線リンクも使用しうる。任意の利用例において、通信インターフェイス 1718 は、様々な種類の情報を表すデジタルデータストリームを担持する電気、電磁または光信号を送受信する。

10

【0097】

ネットワークリンク 1720 は、典型的には、1つ以上のネットワークを介して他のデータ装置へのデータ通信を提供する。例えば、ネットワークリンク 1720 は、ローカルエリアネットワーク 1722 を介してホストコンピュータ 1724 またはインターネットサービスプロバイダ (ISP) 1726 によって運転されるデータ装置への接続を提供しうる。ISP 1726 は、次に、データ通信サービスを、今では一般的に「インターネット」1728 と称されるワールドワイドパケットデータ通信ネットワークを提供する。ローカルエリアネットワーク 1722 およびインターネット 1728 の両方は、デジタルデータストリームを担持する電気、電磁または光信号を用いる。様々なネットワークを通る信号、並びに、ネットワークリンク 1720 上、および、通信インターフェイス 1718 を通る信号は、送信媒体の形態の例であり、デジタルデータをコンピュータシステム 1700 へと、および、そこから搬送する。

20

【0098】

コンピュータシステム 1700 は、ネットワーク、ネットワークリンク 1720 および通信インターフェイス 1718 を介して、メッセージを送信し、プログラムコードを含むデータを受信しうる。インターネットの例において、サーバ 1730 は、要求されたアプリケーションプログラムのコードを、インターネット 1728、ISP 1726、ローカルエリアネットワーク 1722、および、通信インターフェイス 1718 を介して送信しうる。

30

【0099】

受信したコードは、受信された時に、プロセッサ 1704 によって実行されるか、および/または、後で実行されるために他の不揮発性記憶装置に記憶されうる。

【0100】

ここまでの明細書の記載において、実施形態を、利用例により異なりうる多数の具体的な詳細事項について記載した。したがって、本明細書の記載および図面は、限定するものではなく、例示するものであると見なされるべきである。本開示の範囲を唯一かつ排他的に示すもの、および、出願人が本開示の範囲であると意図するものは、本出願から特定の形態で許可される請求項の文言および等価範囲であり、許可請求項は、任意の補正事項を含むものである。

40

【符号の説明】

【0101】

- 1700 コンピュータシステム
- 1704 プロセッサ
- 1712 表示部
- 1714 入力装置

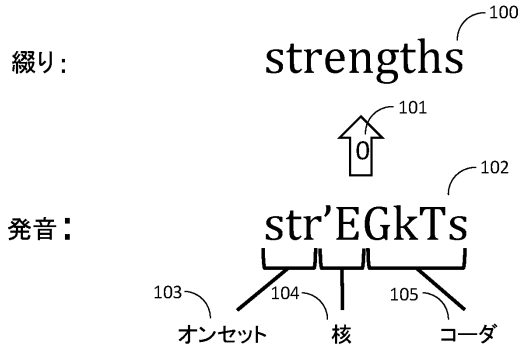
50



1720 ネットワークリンク

【図面】

【図1】



【図2】

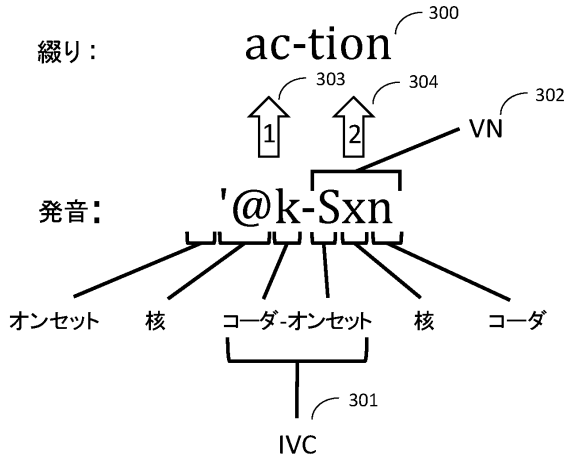
子音 <sup>200</sup>		核 <sup>201</sup>	
IPA	Klattese	IPA	Klattese
p	p	i	i
t	t	l	l
k	k	ε	E
b	b	e	e
d	d	æ	@
g	g	a	a
ʃ	C	au	W
dʒ	J	al	J
s	s	Λ	^
ʃ	S	ɔ	c
z	z	Ol	O
ʒ	Z	o	o
f	f	u	u
θ	T	u	U
v	v	ɜ	R
ð	D	ə	x
h	h	i	
n	n	ɔ	X
m	m	η	N
ŋ	G	η	M
l	l	!	L
J	r	#	切れ目
w	w	'	強勢
j	y		

203

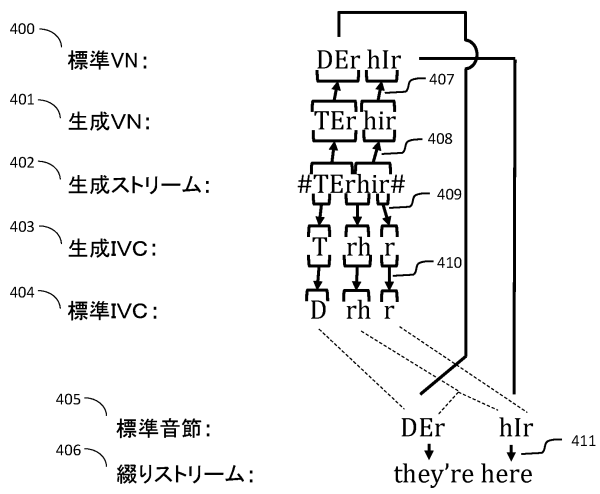
10

20

【図3】



【図4】

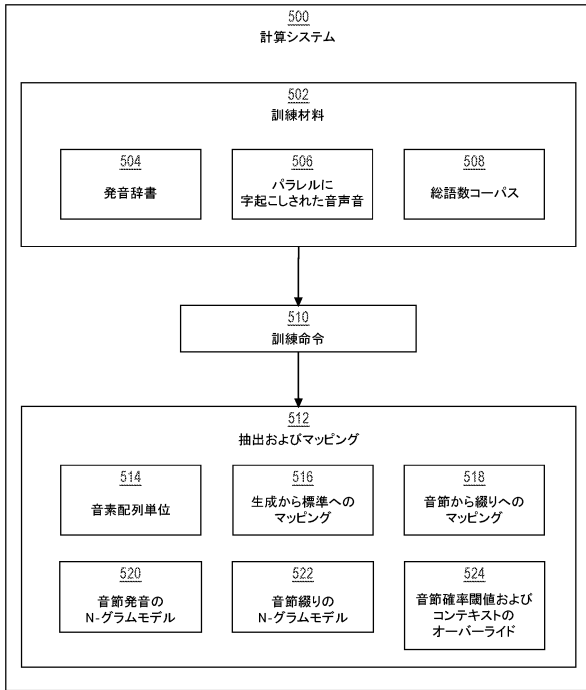


30

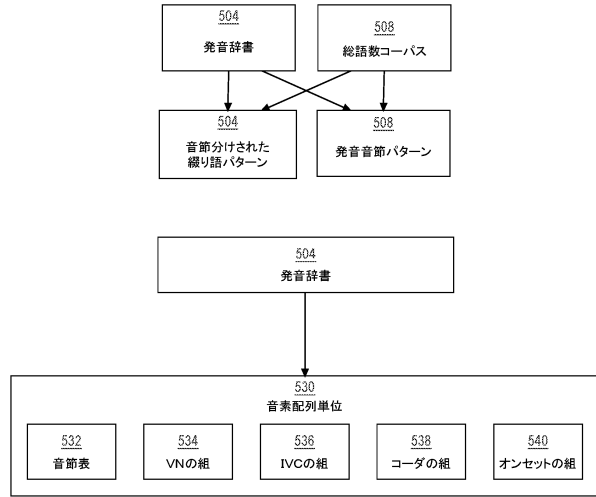
40

50

【図 5 A】



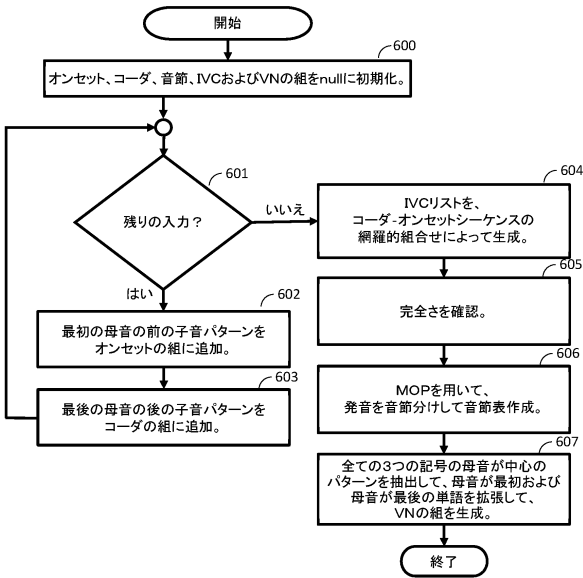
【図 5 B】



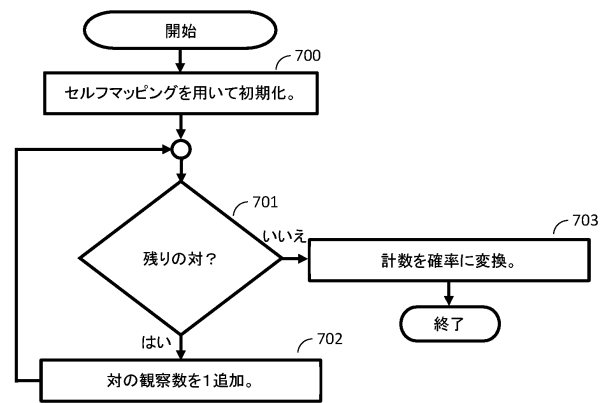
10

20

【図 6】



【図 7】

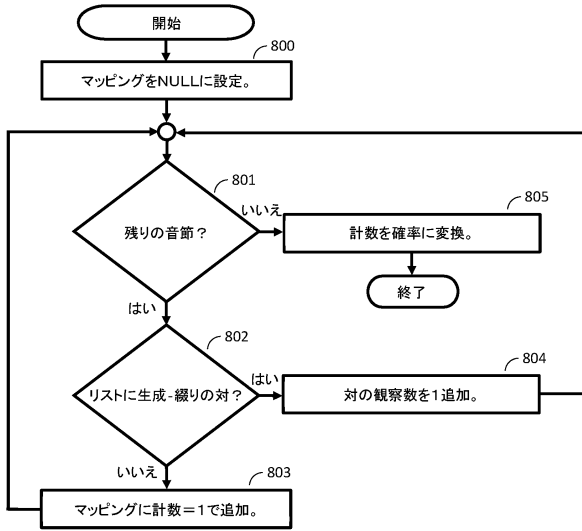


30

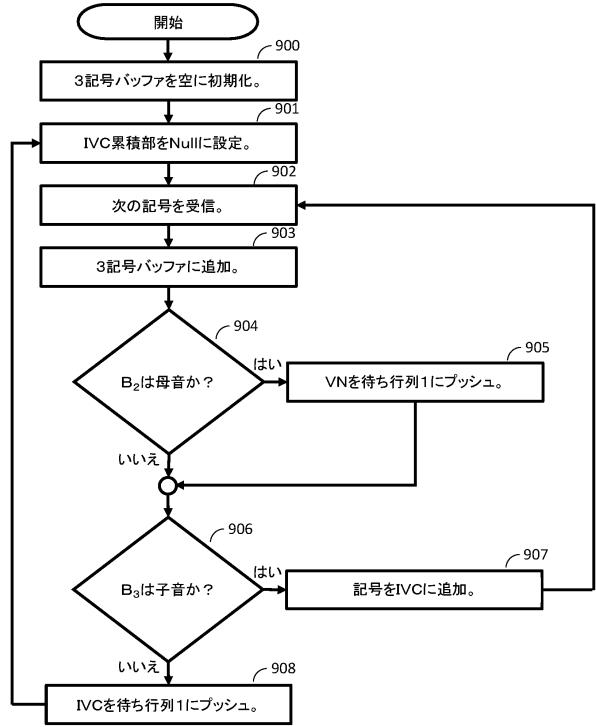
40

50

【図8】



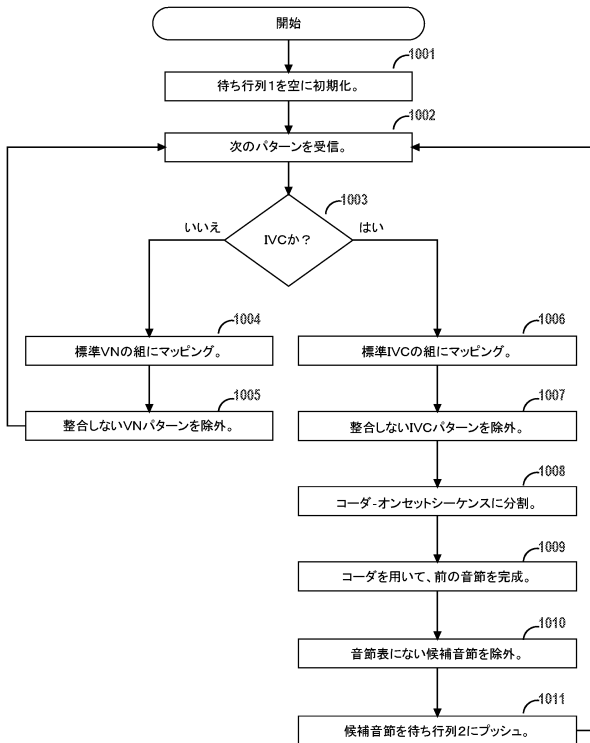
【図9】



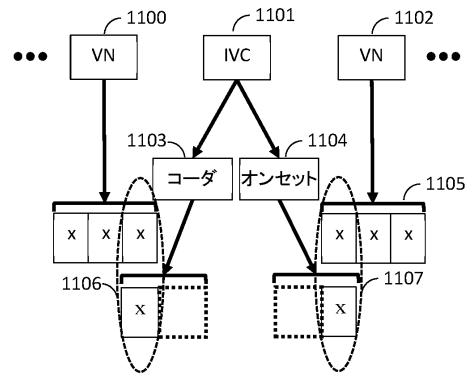
10

20

【図10】



【図11】



30

40

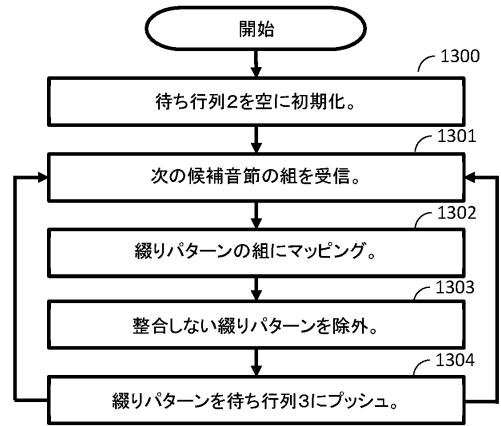
50

【図 1 2】

発音 IVC = /T/

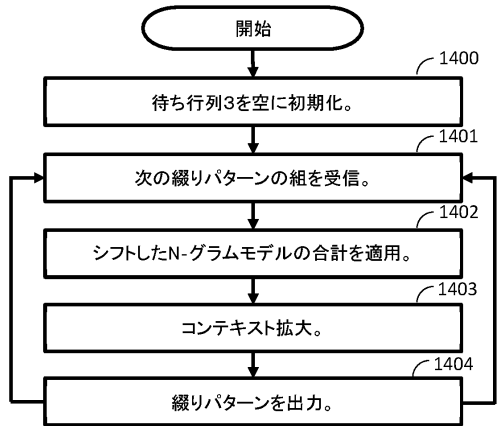
正規化前		正規化後(推定)	
1200	1201	1202	1203
T	0.95219800	T	0.95219800
TD	0.01747250	D	0.00344322
rT	0.00377289	Tr	0.00340659
tT	0.00373626	s	0.00047619
D	0.00344322	●	●
Tr	0.00340659	●	●
TT	0.00318681	●	●
Th	0.00245421	d	3.663e-05
vT	0.00161172	v	3.663e-05
●	●	st	3.663e-05
●	●		
tsT	3.663e-05		
rTD	3.663e-05		
rTh	3.663e-05		

【図 1 3】



10

【図 1 4】



【図 1 5】

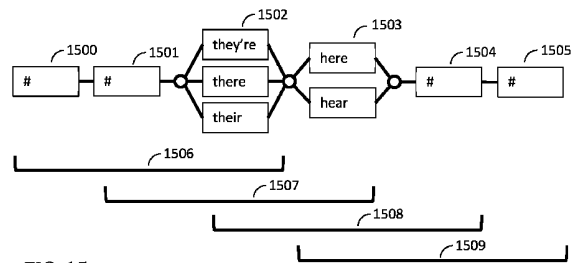


FIG. 15

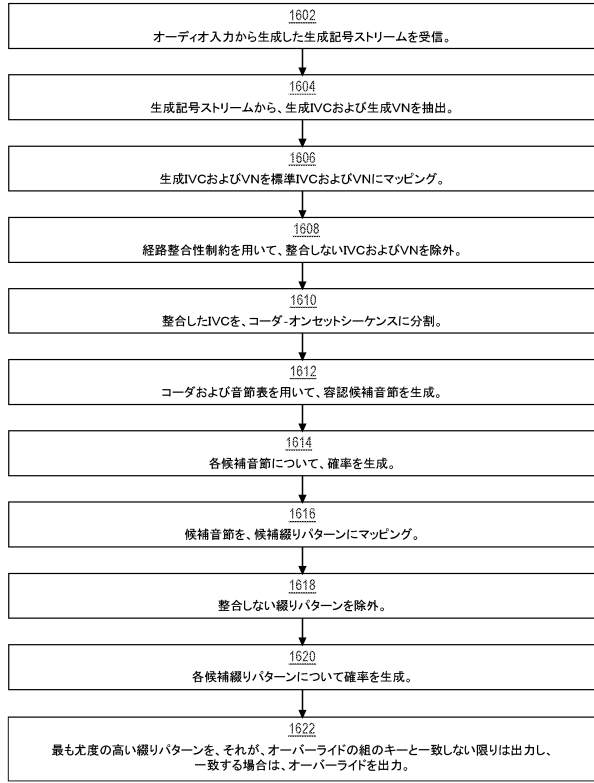
20

30

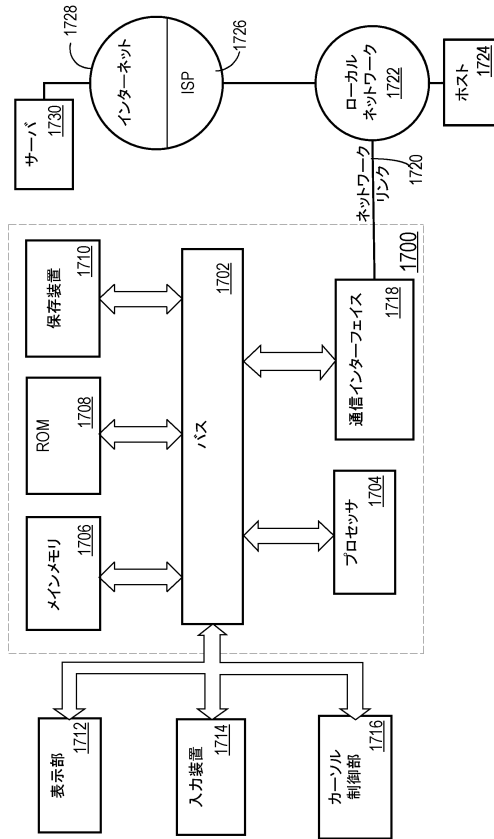
40

50

【図 16】



【図 17】



10

20

30

40

50

---

フロントページの続き

(72)発明者 ピンソン, マーク ビー  
アメリカ合衆国 メリーランド州 20910 シルヴァー スプリング フェントン ストリート 8  
630 ナンバー520

審査官 大野 弘

(56)参考文献 特開2001-195087(JP, A)

(58)調査した分野 (Int.Cl., DB名)

G10L 15/18

G10L 15/10

G10L 15/197