

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7611368号
(P7611368)

(45)発行日 令和7年1月9日(2025.1.9)

(24)登録日 令和6年12月25日(2024.12.25)

(51)国際特許分類		F I	
G 1 0 L	21/12 (2013.01)	G 1 0 L	21/12
G 1 0 L	21/14 (2013.01)	G 1 0 L	21/14
G 1 0 L	25/48 (2013.01)	G 1 0 L	25/48 1 0 0
G 0 9 B	19/06 (2006.01)	G 0 9 B	19/06
請求項の数 34 (全41頁)			
(21)出願番号	特願2023-512331(P2023-512331)	(73)特許権者	513127755 ソムニック インク . アメリカ合衆国 カリフォルニア サニー ヴェール サウス カシミア テラス 5 0 9
(86)(22)出願日	令和3年8月17日(2021.8.17)	(74)代理人	100114890 弁理士 アインゼル・フェリックス＝ラ インハルト
(65)公表番号	特表2023-539148(P2023-539148 A)	(74)代理人	100098501 弁理士 森田 拓
(43)公表日	令和5年9月13日(2023.9.13)	(74)代理人	100116403 弁理士 前川 純一
(86)国際出願番号	PCT/US2021/046366	(74)代理人	100134315 弁理士 永島 秀郎
(87)国際公開番号	WO2022/040229	(74)代理人	100162880 弁理士 上島 類
(87)国際公開日	令和4年2月24日(2022.2.24)		
審査請求日	令和5年5月22日(2023.5.22)		
(31)優先権主張番号	63/068,734		
(32)優先日	令和2年8月21日(2020.8.21)		
(33)優先権主張国・地域又は機関	米国(US)		
		最終頁に続く	

(54)【発明の名称】 発話をコンピュータ生成によって視覚化するための方法およびシステム

(57)【特許請求の範囲】

【請求項 1】

少なくとも1つの分節を含む発話をコンピュータ生成によって視覚化する方法であって、前記方法は、

前記発話の分節に対応するオブジェクトの図像表現を生成することであって、前記図像表現を生成することは、

前記分節の持続時間を、前記オブジェクトの長さによって表現することと、
前記分節の強さを、前記オブジェクトの幅によって表現することと、
前記分節のピッチ曲線を、基準フレームに対する前記オブジェクトの傾斜角によって表現することと
を含む、ことと、

前記オブジェクトの図像表現を、コンピューティング装置の画面上に表示することとを含む、方法。

【請求項 2】

前記ピッチ曲線は、基本周波数の動きに関連付けられており、
前記図像表現を生成することは、前記分節の前記基本周波数のオフセットを、前記基準フレームに対する前記オブジェクトの垂直方向の位置によって表現することをさらに含む、請求項1記載の方法。

【請求項 3】

前記分節は、第1の分節であり、

前記方法は、

前記第 1 の分節に対応する第 1 のオブジェクトを表示することと、

前記発話の、前記第 1 の分節に後続する第 2 の分節に対応する第 2 のオブジェクトを、前記第 1 のオブジェクトと前記第 2 のオブジェクトとが、前記第 1 の分節と前記第 2 の分節との間の無声音期間に対応する間隔によって分離されるように表示することを含む、請求項 1 記載の方法。

【請求項 4】

少なくとも 1 つの分節を含む発話をコンピュータ生成によって視覚化する方法であって、前記方法は、

複数のオブジェクトを含む図像表現を生成することであって、前記複数のオブジェクトの各々は、前記発話のそれぞれの分節に対応し、前記発話の前記分節は、子音、母音またはそれらの組み合わせを含み、各オブジェクトの前記図像表現は、前記発話の前記それぞれの分節の発音に基づいており、前記図像表現を生成することは、前記複数のオブジェクトの各々ごとに、

前記それぞれの分節の持続時間を、前記オブジェクトの長さによって表現すること、および前記それぞれの分節の強さを、前記オブジェクトの幅によって表現することと、

前記図像表現において、隣接するオブジェクト同士の間隔を配置することを含む、ことと、

前記図像表現を、コンピューティング装置の画面上に表示することと、

第 1 の話者によって発話された、前記発話の複数の分節のオブジェクトの第 1 の集合の第 1 の視覚化部を生成し、前記画面上に表現することと、

第 2 の話者によって発話された、前記発話の複数の分節のオブジェクトの第 2 の集合の第 2 の視覚化部を生成し、前記画面上に表現することとを含み、オブジェクトの前記第 1 の集合またはオブジェクトの前記第 2 の集合は、前記オブジェクトを含む、

方法。

【請求項 5】

前記複数のオブジェクトの各々は、境界線によって画定され、

前記図像表現における 2 つの隣接するオブジェクトの境界線同士の間隔の前記間隔は、無声音期間の持続時間に基づいている、

請求項 4 記載の方法。

【請求項 6】

少なくとも 1 つの分節を含む発話をコンピュータ生成によって視覚化する方法であって、前記方法は、

複数のオブジェクトを含む図像表現を生成することであって、前記複数のオブジェクトの各々は、前記発話のそれぞれの分節に対応し、前記図像表現を生成することは、前記複数のオブジェクトの各々ごとに、

前記それぞれの分節の持続時間を、前記オブジェクトの長さによって表現すること、および前記それぞれの分節の強さを、前記オブジェクトの幅によって表現することと、

前記図像表現において、隣接するオブジェクト同士の間隔を配置することとを含む、ことと、

前記図像表現を、コンピューティング装置の画面上に表示することと、

前記オブジェクトを、前記分節に対応する音の調音部位および / または調音方法に基づいて選択された色で表示することと

を含む、方法。

【請求項 7】

前記分節は、少なくとも 1 つの音素を含む、

請求項 4 記載の方法。

【請求項 8】

前記分節は、前記少なくとも 1 つの音素に少なくとも 1 つの母音を含む、

請求項 7 記載の方法。

【請求項 9】

少なくとも 1 つの分節を含む発話をコンピュータ生成によって視覚化する方法であって、
前記方法は、
複数のオブジェクトを含む図像表現を生成することであって、前記複数のオブジェクトの
各々は、前記発話のそれぞれの分節に対応し、前記図像表現を生成することは、前記複数
のオブジェクトの各々ごとに、
前記それぞれの分節の持続時間を、前記オブジェクトの長さによって表現すること、およ
び前記それぞれの分節の強さを、前記オブジェクトの幅によって表現することと、
前記図像表現において、隣接するオブジェクト同士の間には間隔を配置することと
を含む、ことと、
前記図像表現を、コンピューティング装置の画面上に表示することと、
前記オブジェクトを、前記分節における最初の音素に基づいて選択された色で表示するこ
とと
を含み、前記分節は、少なくとも 1 つの音素を含む、
方法。

10

【請求項 10】

前記方法は、
前記発話を、少なくとも 1 つの音素を含む分節に分解することと、
前記少なくとも 1 つの音素を、前記オブジェクトに付随する少なくとも 1 つの記号とし
て表示することと
をさらに含む、請求項 7 記載の方法。

20

【請求項 11】

前記第 1 の視覚化部は、
第 1 の話者によって発話された第 1 の発話を表現し、前記第 1 の視覚化部は、前記第 1
の発話に対応するオブジェクトの第 1 の集合を前記画面上に含み、
前記第 2 の視覚化部は、
第 2 の話者によって発話された第 2 の発話を表現し、前記第 2 の視覚化部は、前記第 2
の発話に対応するオブジェクトの第 2 の集合を含み、前記第 1 の集合または前記第 2 の集
合は、前記オブジェクトを含み、
前記第 2 の視覚化部は、オブジェクトの前記第 1 の集合の第 1 の端部と、オブジェクトの
前記第 2 の集合の第 1 の端部とが前記画面上で実質的に垂直方向に整列するように、前記
画面上に表示される、請求項 4 記載の方法。

30

【請求項 12】

前記コンピューティング装置は、マイクロフォン入力部をさらに含み、
前記方法は、
前記第 1 の視覚化部を表示することに続いて、前記マイクロフォン入力部を介して前記
第 2 の発話を録音することと、
録音された前記第 2 の発話に回答して、前記第 2 の視覚化部を生成して表示することと
をさらに含む、請求項 11 記載の方法。

【請求項 13】

前記オブジェクトは、長方形、楕円形、およびたまご形から選択される形状を有する、
請求項 4 記載の方法。

40

【請求項 14】

少なくとも 1 つの分節を含む発話をコンピュータ生成によって視覚化する方法であって、
前記方法は、
複数のオブジェクトを含む図像表現を生成することであって、前記複数のオブジェクトの
各々は、前記発話のそれぞれの分節に対応し、前記図像表現を生成することは、前記複数
のオブジェクトの各々ごとに、
前記それぞれの分節の持続時間を、前記オブジェクトの長さによって表現すること、およ
び前記それぞれの分節の強さを、前記オブジェクトの幅によって表現することと、

50

前記図像表現において、隣接するオブジェクト同士の間隔を配置することとを含む、ことと、

前記図像表現を、コンピューティング装置の画面上に表示することとを含む、

前記オブジェクトの傾斜角は、前記オブジェクトの長さによって変化する、方法。

【請求項 15】

非一時的コンピュータ可読媒体であって、

請求項 1 から 14 までのいずれか 1 項記載の方法を実施するためにコンピューティング装置によって実行可能である命令が、前記非一時的コンピュータ可読媒体上に保存されている、非一時的コンピュータ可読媒体。

10

【請求項 16】

コンピューティング装置と、請求項 15 記載の非一時的コンピュータ可読媒体とを含む、システム。

【請求項 17】

前記コンピューティング装置は、前記非一時的コンピュータ可読媒体を含むメモリを含む、

請求項 16 記載のシステム。

【請求項 18】

非一時的コンピュータ可読媒体であって、

20

以下のオペレーションを実施するためにコンピューティング装置によって実行可能である命令が、前記非一時的コンピュータ可読媒体上に保存されており、

前記オペレーションは、

発話の視覚化部を生成することであって、前記視覚化部は、前記発話の分節に対応するオブジェクトを含み、前記視覚化部を生成することは、

前記分節の持続時間を、前記オブジェクトの長さによって表現することと、

前記分節の強さを、前記オブジェクトの幅によって表現することと、

前記分節のピッチ曲線を、基準フレームに対する前記オブジェクトの傾斜角によって表現することと

を含む、ことと、

30

前記視覚化部を、前記コンピューティング装置に結合された画面上に表示することとを含む、非一時的コンピュータ可読媒体。

【請求項 19】

前記オブジェクトは、規則的な幾何形状を有する 2 次元のオブジェクトである、請求項 18 記載の非一時的コンピュータ可読媒体。

【請求項 20】

前記オブジェクトは、たまご形、楕円形、および長方形から選択される形状を有する、請求項 18 記載の非一時的コンピュータ可読媒体。

【請求項 21】

前記ピッチ曲線は、基本周波数の動きに関連付けられており、

40

前記視覚化部を生成することは、前記分節の前記基本周波数のオフセットを、前記基準フレームに対する前記オブジェクトの垂直方向の位置によって表現することをさらに含む、請求項 18 記載の非一時的コンピュータ可読媒体。

【請求項 22】

前記分節は、前記発話の第 1 の分節であり、

前記オペレーションは、

前記第 1 の分節に対応する第 1 のオブジェクトを表示することと、

前記発話の、前記第 1 の分節に後続する第 2 の分節に対応する第 2 のオブジェクトを表示することと

を含む、

50

前記第 1 のオブジェクトと前記第 2 のオブジェクトとは、前記第 1 の分節と前記分節との間の無声音期間に対応する間隔によって分離される、
請求項 18 記載の非一時的コンピュータ可読媒体。

【請求項 23】

前記分節は、少なくとも 1 つの音素を含む、
請求項 18 記載の非一時的コンピュータ可読媒体。

【請求項 24】

前記分節は、前記少なくとも 1 つの音素に少なくとも 1 つの母音を含む、
請求項 23 記載の非一時的コンピュータ可読媒体。

【請求項 25】

前記オペレーションは、前記オブジェクトを、前記分節における最初の音素に基づいて選択された色で表示することをさらに含む、
請求項 23 記載の非一時的コンピュータ可読媒体。

【請求項 26】

前記色は、前記分節に対応する音の調音部位および／または調音方法に基づいて選択される、
請求項 25 記載の非一時的コンピュータ可読媒体。

【請求項 27】

前記オペレーションは、
前記発話を、少なくとも 1 つの音素を含む少なくとも 1 つの分節に分解することと、
前記少なくとも 1 つの音素を、前記オブジェクトと一緒に前記視覚化部における対応する数の記号として表現することと
をさらに含む、請求項 23 記載の非一時的コンピュータ可読媒体。

【請求項 28】

前記オペレーションは、
第 1 の話者によって発話された第 1 の発話の第 1 の視覚化部を生成して前記画面上に表示することであって、前記第 1 の視覚化部は、前記第 1 の発話に対応するオブジェクトの第 1 の集合を前記画面上に含む、ことと、

第 2 の話者によって発話された第 2 の発話の第 2 の視覚化部を生成することであって、
前記第 2 の視覚化部は、前記第 2 の発話に対応するオブジェクトの第 2 の集合を含む、ことと、

前記オブジェクトの前記第 1 の集合の第 1 の端部と、前記オブジェクトの前記第 2 の集合の第 1 の端部とが前記画面上で実質的に垂直方向に整列するように、前記第 2 の視覚化部を前記画面上に表示することと

を含む、請求項 18 記載の非一時的コンピュータ可読媒体。

【請求項 29】

前記コンピューティング装置は、マイクロフォン入力部に結合されており、
前記オペレーションは、
前記第 1 の視覚化部を表示することについて、前記マイクロフォン入力部を介して前記第 2 の発話を録音することと、

録音された前記第 2 の発話に回答して、前記第 2 の視覚化部を生成して表示することと
をさらに含む、請求項 28 記載の非一時的コンピュータ可読媒体。

【請求項 30】

前記コンピューティング装置は、音響出力部に結合されており、
前記オペレーションは、
前記音響出力部を介して前記第 1 の発話の音響再生を提供することと、
前記第 2 の視覚化部を表示することについて、ユーザが前記第 1 の発話の前記音響再生を再生することを可能にするように構成されたユーザコントロールを提供することと
をさらに含む、請求項 28 記載の非一時的コンピュータ可読媒体。

【請求項 31】

10

20

30

40

50

システムであって、前記システムは、
プロセッサと、
ディスプレイと、

前記プロセッサによって実行された場合に、前記プロセッサに以下のオペレーションを実施させるための命令を含むメモリと

を含み、前記オペレーションは、

発話の分節に対応するオブジェクトの画像表現を生成することであって、前記画像表現は、

前記分節の持続時間を、前記オブジェクトの長さによって表現することと、

前記分節の強さを、前記オブジェクトの幅によって表現することと、

前記分節のピッチ曲線を、基準フレームに対する前記オブジェクトの傾斜角によって表現することと

によって生成される、ことと、

前記オブジェクトの画像表現を、コンピューティング装置の画面上に表示することとを含む、システム。

【請求項 3 2】

前記分節は第 1 の分節であり、

前記オペレーションは、

前記第 1 の分節に対応する第 1 のオブジェクトを表示することと、

前記発話の、前記第 1 の分節に後続する第 2 の分節に対応する第 2 のオブジェクトを表示することと、

前記第 1 のオブジェクトと前記第 2 のオブジェクトとの間に、前記第 1 の分節と前記第 2 の分節との間の無声音期間に対応する間隔を配置することと

をさらに含む、請求項 3 1 記載のシステム。

【請求項 3 3】

前記オペレーションは、前記オブジェクトを、前記分節に対応する音の調音部位および / または調音方法に基づいて選択された色で表示することをさらに含む、
請求項 3 1 記載のシステム。

【請求項 3 4】

前記オペレーションは、

第 1 の話者によって発話された第 1 の発話の第 1 の視覚化部を生成して前記画面上に表示することであって、前記第 1 の視覚化部は、前記第 1 の発話に対応するオブジェクトの第 1 の集合を前記画面上に含む、ことと、

第 2 の話者によって発話された第 2 の発話の第 2 の視覚化部を生成することであって、前記第 2 の視覚化部は、前記第 2 の発話に対応するオブジェクトの第 2 の集合を含む、ことと、

前記オブジェクトの前記第 1 の集合の第 1 の端部と、前記オブジェクトの前記第 2 の集合の第 1 の端部とが前記画面上で実質的に垂直方向に整列するように、前記第 2 の視覚化部を前記画面上に表示することと

をさらに含む、請求項 3 1 記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

本出願は、2020年8月21日に提出された米国仮出願第63/068,734号明細書の優先権を主張するものであり、如何なる目的であってもその全てが参照により本明細書に援用されるものとする。

【0002】

技術分野

本発明は、一般的には、発話による言語学習のための方法、システム、および装置に関し、より詳細には、言語学習者のための、発話をコンピュータ生成によって視覚化する方

10

20

30

40

50

法およびシステムに関する。

【 0 0 0 3 】

背景

人間は、発声された表出、典型的には発話によって情報を伝達する。人間が発話を生成している間に伝達される情報は、言語情報と、パラ言語情報と、非言語情報とに分類可能である。言語情報は、一般的に書記された形態で表現される。パラ言語情報は、発話中に言語情報に伴い得る。非言語情報は、発話中に伝達される言語情報から独立し得る。

【 0 0 0 4 】

例えば英語の場合には、言語情報は、ローマ字アルファベットの文字列で表現することができる音素特徴に関連付けられている。音素とは、英語における子音および母音のような、特定の言語における音の知覚的に異なる単位である。英語においてそれぞれの音素を表現する際には、1つまたは2つのローマ字アルファベットを使用することができる。アルファベットの文字列は、1つまたは複数の音節を含むことができる1つの単語を構成し、この場合、それぞれの音節は、典型的には1つの母音を含み、母音を取り囲む1つまたは複数の子音も含むことができる。母音は、例えば、聴者の母音の知覚を主に支配する比較的低いフォルマント周波数（例えば、 F_1 および F_2 ）のような物理的なパラメータによって観察され得る。フォルマント周波数は、スペクトログラム上の局所的な最大値として取得される。フォルマント周波数は、人間の声道の音響共鳴を表現することが知られている。子音は、非周期的な信号として観測され得るか、またはスペクトログラムの高周波領域における周期的な信号として観測され得る。英語におけるパラ言語情報は、通常、韻律特徴によって表現される。例えば、韻律特徴は、ストレス、リズム、およびピッチを含む。ストレスは、強さとして観測され得る。リズムは、それぞれの音素または音節の持続時間と、音素同士または音節同士の間の休止とを含む時間的なパラメータである。ピッチは、発話を伝達する音声の知覚される高さである。ピッチは、スペクトログラム上の基本周波数（例えば、 F_0 ）として観察され得る。

【 0 0 0 5 】

発話の従来の視覚的表現は、時間軸および周波数軸によって画定される平面上の濃淡として強さを示すスペクトログラムと、国際音声記号（International Phonetic Alphabets: IPA）のような音声表記を伴う、抽出された音響パラメータ（例えば、 F_0 、 F_1 、および F_2 ）の曲線とに、大きく依存してきた。IPAのそれぞれのアルファベットは、それぞれの音素に対応しており、IPAを用いると、“right”および“write”のようなIPAによって同一に表現される可能性のあるバリエーションを有する、英語のローマ字アルファベットを使用したテキスト表現に関係なく、音素の発音が正確に表現されるという利点がある。

【 0 0 0 6 】

しかしながら、発話のこのような従来の視覚的表現、すなわちスペクトログラム表現およびIPA表記は、ユーザにとって直感的ではなく、またユーザフレンドリーでもなかった。ユーザが（例えば、ネイティブスピーカーおよび熟練した第二言語教師によって提供される）基準発話の録音と自身の発話の録音との間の違いを、発話の視覚的表現を介して直感的に学習することが可能となるように、発話のよりユーザフレンドリーな視覚的表現が望まれている。

【 0 0 0 7 】

概要

少なくとも1つの分節を含む図像表現のためのシステムおよび方法が説明されている。いくつかの実施形態によれば、少なくとも1つの分節を含む発話をコンピュータ生成によって視覚化する方法は、発話の分節に対応するオブジェクトの図像表現を生成することであって、図像表現を生成することは、少なくとも、分節の持続時間を、オブジェクトの長さによって表現することと、分節の強さを、オブジェクトの幅によって表現することと、分節のピッチ曲線を、基準フレームに対するオブジェクトの傾斜角によって表現することとを含む、ことを含み、その後、オブジェクトの図像表現が、コンピューティング装置の

10

20

30

40

50

画面上に表示される。ピッチ曲線が、基本周波数の動きに関連付けられているいくつかの実施形態では、図像表現を生成することは、分節の基本周波数のオフセットを、基準フレームに対するオブジェクトの垂直方向の位置によって表現することをさらに含む。いくつかの実施形態では、分節は、第1の分節であり、本方法は、第1の分節に対応する第1のオブジェクトを表示することと、発話の、第1の分節に後続する第2の分節に対応する第2のオブジェクトを、第1のオブジェクトと第2のオブジェクトとが、第1の分節と第2の分節との間の無声音期間に対応する間隔によって分離されるように表示することとを含む。いくつかの実施形態では、本方法は、複数のオブジェクトを含む図像表現を生成することと、複数のオブジェクトの各々は、発話のそれぞれの分節に対応し、図像表現を生成することは、複数のオブジェクトの各々ごとに、それぞれの分節の持続時間を、オブジェクトの長さによって表現すること、およびそれぞれの分節の強さを、オブジェクトの幅によって表現することと、図像表現において、隣接するオブジェクト同士の間隔を配置することとを含む、こととを含む。いくつかの実施形態では、複数のオブジェクトの各々は、境界線によって画定され、図像表現における2つの隣接するオブジェクトの境界線同士の間隔は、無声音期間の持続時間に基づいている。いくつかの実施形態では、本方法は、オブジェクトを、分節に対応する音の調音部位および/または調音方法に基づいて選択された色で表示することをさらに含む。いくつかの実施形態では、分節は、少なくとも1つの音素を含む。いくつかの実施形態では、分節は、少なくとも1つの音素に少なくとも1つの母音を含む。いくつかの実施形態では、本方法は、オブジェクトを、分節における最初の音素に基づいて選択された色で表示することを含む。いくつかの実施形態では、本方法は、発話を、少なくとも1つの音素を含む分節に分解することと、少なくとも1つの音素を、オブジェクトに付随する少なくとも1つの記号として表示することとを含む。いくつかの実施形態では、本方法は、第1の話者によって発話された第1の発話の第1の視覚化部を生成して画面上に表示することと、第1の視覚化部は、第1の発話に対応するオブジェクトの第1の集合を画面上に含む、ことと、第2の話者によって発話された第2の発話の第2の視覚化部を生成することと、第2の視覚化部は、第2の発話に対応するオブジェクトの第2の集合を含む、ことと、オブジェクトの第1の集合の第1の端部と、オブジェクトの第2の集合の第1の端部とが画面上で実質的に垂直方向に整列するように、第2の視覚化部を画面上に表示することとを含む。コンピューティング装置が、マイクロフォン入力部を含んでいる本方法のいくつかの実施形態では、本方法は、第1の視覚化部を表示することに続いて、マイクロフォン入力部を介して第2の発話を録音することと、録音された第2の発話に回答して、第2の視覚化部を生成して表示することとを含む。いくつかの実施形態では、オブジェクトは、長方形、楕円形、およびたまご形から選択される形状を有する。いくつかの実施形態では、オブジェクトの傾斜角は、オブジェクトの長さに沿って変化する。

【0008】

本明細書では、コンピューティング装置の1つまたは複数のプロセッサによって実行された場合に、本明細書の任意の例による方法をコンピューティング装置に実施させるための命令を有する、非一時的コンピュータ可読媒体の実施形態が開示されている。本明細書の任意の実施例による非一時的コンピュータ可読媒体は、コンピューティングシステムの一部であってよく、コンピューティングシステムは、オプションとしてディスプレイを含むことができる。いくつかの実施形態では、非一時的コンピュータ可読媒体を、発話のコンピュータ生成による視覚化部を表示するコンピューティング装置のメモリによって提供することができる。

【0009】

いくつかの実施形態では、発話の視覚化部を生成するためにコンピューティング装置によって実行可能である命令が、非一時的コンピュータ可読媒体上に保存されており、視覚化部は、発話の分節に対応するオブジェクトを含む。いくつかの実施形態では、発話の視覚化部を生成することは、分節の持続時間を、オブジェクトの長さによって表現することと、分節の強さを、オブジェクトの幅によって表現することと、分節のピッチ曲線を、基

10

20

30

40

50

準フレームに対するオブジェクトの傾斜角によって表現することを含む。命令はさらに、視覚化部を、コンピューティング装置に結合された画面上に表示することをコンピューティング装置に実施させる。いくつかの実施形態では、オブジェクトは、規則的な幾何形状を有する２次元のオブジェクトである。いくつかの実施形態では、オブジェクトは、たまご形、楕円形、および長方形から選択される形状を有する。ピッチ曲線が、基本周波数の動きに関連付けられているいくつかの実施形態では、視覚化部を生成することは、分節の基本周波数のオフセットを、基準フレームに対するオブジェクトの垂直方向の位置によって表現することをさらに含む。分節が、発話の第１の分節であるいくつかの実施形態では、命令はさらに、第１の分節に対応する第１のオブジェクトを表示することと、発話の、第１の分節に後続する第２の分節に対応する第２のオブジェクトを表示することとをコンピューティング装置に実施させ、第１のオブジェクトと第２のオブジェクトとは、第１の分節と第２の分節との間の無声音期間に対応する間隔によって分離される。いくつかの実施形態では、分節は、少なくとも１つの音素を含む。いくつかの実施形態では、分節は、少なくとも１つの音素に少なくとも１つの母音を含む。いくつかの実施形態では、命令はさらに、オブジェクトを、分節における最初の音素に基づいて選択された色で表示することをコンピューティング装置に実施させる。いくつかの実施形態では、色は、分節に対応する音の調音部位および／または調音方法に基づいて選択される。いくつかの実施形態では、命令はさらに、発話を、少なくとも１つの音素を含む少なくとも１つの分節に分解することと、少なくとも１つの音素を、オブジェクトと一緒に視覚化部における対応する数の記号として表現することとをコンピューティング装置に実施させる。いくつかの実施形態では、命令はさらに、第１の話者によって発話された第１の発話の第１の視覚化部を生成して画面上に表示することと、第１の視覚化部は、第１の発話に対応するオブジェクトの第１の集合を画面上に含む、ことと、第２の話者によって発話された第２の発話の第２の視覚化部を生成することと、第２の視覚化部は、第２の発話に対応するオブジェクトの第２の集合を含む、ことと、オブジェクトの第１の集合の第１の端部と、オブジェクトの第２の集合の第１の端部とが画面上で実質的に垂直方向に整列するように、第２の視覚化部を画面上に表示することとをコンピューティング装置に実施させる。コンピューティング装置が、マイクロフォン入力部に結合されているいくつかの実施形態では、命令はさらに、第１の視覚化部を表示することと、マイクロフォン入力部を介して第２の発話を録音することと、録音された第２の発話に回答して、第２の視覚化部を生成して表示することとをコンピューティング装置に実施させる。コンピューティング装置が、音響出力部に結合されているいくつかの実施形態では、命令はさらに、音響出力部を介して第１の発話の音響再生を提供することと、第２の視覚化部を表示することと、ユーザが第１の発話の音響再生を再生することを可能にするように構成されたユーザコントロールを提供することとをコンピューティング装置に実施させる。

【００１０】

本明細書のいくつかの実施形態によるシステムは、プロセッサと、ディスプレイと、プロセッサによって実行された場合に、本明細書で説明されている発話の視覚化部を生成することに関連するオペレーションのいずれかをプロセッサに実施させるための命令を含むメモリとを含む。いくつかの実施形態では、これらのオペレーションは、第１の分節に対応する第１のオブジェクトを表示することと、発話の、第１の分節に後続する第２の分節に対応する第２のオブジェクトを表示することと、第１のオブジェクトと第２のオブジェクトとの間に、第１の分節と第２の分節との間の無声音期間に対応する間隔を配置することとを含む。いくつかの実施形態では、オペレーションは、オブジェクトを、分節に対応する音の調音部位および／または調音方法に基づいて選択された色で表示することをさらに含む。いくつかの実施形態では、オペレーションは、第１の話者によって発話された第１の発話の第１の視覚化部を生成して画面上に表示することと、第１の視覚化部は、第１の発話に対応するオブジェクトの第１の集合を画面上に含む、ことと、第２の話者によって発話された第２の発話の第２の視覚化部を生成することと、第２の視覚化部は、第２の発話に対応するオブジェクトの第２の集合を含む、ことと、オブジェクトの

10

20

30

40

50

第 1 の集合の第 1 の端部と、オブジェクトの第 2 の集合の第 1 の端部とが画面上で実質的に垂直方向に整列するように、第 2 の視覚化部を画面上に表示することとをさらに含む。本明細書における発明の主題は、この概要セクションで概説された実施形態に限定されているわけではない。

【図面の簡単な説明】

【 0 0 1 1 】

【図 1】本開示の実施形態による装置の簡略化されたブロック図である。

【図 2 A】本開示の実施形態による、発話の分節化プロセスのフロー図である。

【図 2 B】本開示の実施形態による、分節の視覚的表現を生成するフロー図である。

【図 2 C】本開示の実施形態による、発話の生成された視覚的表現のタイミング図である。

10

【図 2 D】発話の視覚的表出または視覚的表現の種々異なるバリエーションを示す図である。

【図 2 E】発話の視覚的表出または視覚的表現の種々異なるバリエーションを示す図である。

【図 2 F】発話の視覚的表出または視覚的表現の種々異なるバリエーションを示す図である。

【図 2 G】発話の視覚的表出または視覚的表現の種々異なるバリエーションを示す図である。

【図 3 A】本開示の実施形態による、分節の視覚的表現を生成するフロー図である。

【図 3 B】本開示の実施形態による、色と、子音を含む音素と、子音に関連する調音部位との関係を示す概略図である。

20

【図 3 C】本開示の実施形態による、発話の生成された視覚的表現のタイミング図である。

【図 3 D】本開示の実施形態による、発話の生成された視覚的表現と、発話に関連する顔表現とを含む画面の概略図である。

【図 4 A】本開示の実施形態による、分節の視覚的表現を生成するフロー図である。

【図 4 B】本開示の実施形態による、波形と、スペクトログラムと、スペクトログラムに重畳された発話の生成された視覚的表現とのタイミング図である。

【図 5 A】本開示の実施形態による、波形と、スペクトログラムと、スペクトログラムに重畳された発話の生成された視覚的表現とのタイミング図である。

【図 5 B】本開示の実施形態による、波形と、スペクトログラムと、スペクトログラムに重畳された発話の生成された視覚的表現とのタイミング図である。

30

【図 6 A】本開示の実施形態による、波形と、スペクトログラムと、スペクトログラムに重畳された発話の生成された視覚的表現とのタイミング図である。

【図 6 B】本開示の実施形態による、発話の生成された視覚的表現の概略図である。

【図 6 C】本開示の実施形態による、発話の生成された視覚的表現の概略図である。

【図 7 A】本開示の実施形態による、波形と、スペクトログラムと、スペクトログラムに重畳された発話の生成された視覚的表現とのタイミング図である。

【図 7 B】本開示の実施形態による、発話の生成された視覚的表現の概略図である。

【図 8】図 8 A ~ C は、本開示の実施形態による、発話の生成された視覚的表現の概略図である。

40

【図 9】本開示の実施形態による、発話の視覚的表現を修正するフローの概略図である。

【図 10】図 10 A ~ D は、本開示の実施形態による、発話の生成された視覚的表現をそのタッチ画面上に含んでいる言語学習システムを提供する装置の概略図である。

【図 11】図 11 A ~ E は、本開示の実施形態による、発話の生成された視覚的表現をそのタッチ画面上に含んでいる言語学習システムを提供する装置の概略図である。

【図 12 A】本開示の実施形態による、発話の生成された視覚的表現をそのタッチ画面上に含んでいるコミュニケーションシステムを提供する装置の概略図である。

【図 12 B】本開示の実施形態による、発話の生成された視覚的表現をそのタッチ画面上に含んでいるコミュニケーションシステムを提供する装置の概略図である。

【図 12 C】本開示の実施形態による、発話の生成された視覚的表現をそのタッチ画面上

50

に含んでいるコミュニケーションシステムを提供する装置の概略図である。

【図 1 2 D】本開示の実施形態による、発話の生成された視覚的表現をそのタッチ画面上に含んでいるコミュニケーションシステムを提供する装置の概略図である。

【 0 0 1 2 】

詳細な説明

以下では、本開示の種々の実施形態について添付の図面を参照しながら詳細に説明する。以下の詳細な説明は、本発明を実施することができる特定の態様および実施形態を例示的に示している添付の図面を参照する。これらの実施形態は、当業者が本発明を実施することを可能にするために十分に詳細に説明されている。他の実施形態を利用してもよく、本発明の範囲から逸脱することなくアルゴリズム、構造、およびロジックの変更を行ってもよい。いくつかの開示されている実施形態を、1つまたは複数の他の開示されている実施形態と組み合わせて新しい実施形態を形成することができるので、本明細書に開示されている種々の実施形態は、必ずしも相互に排他的であるとは限らない。

【 0 0 1 3 】

本開示によれば、発話のコンピュータ生成による視覚化部を提供するための装置、システム、および方法が開示されている。いくつかの実施形態では、（例えば、録音された発話から）検出して、現在公知のまたは後々開発される発話認識技術を介して処理することができる発話は、複数の分節を含むことができ、したがって、複数の分節に分節化され得る。いくつかの実施形態では、1つまたは複数の個々の分節は、少なくとも1つの音素を含むことができる。いくつかの実施形態では、分節は、音節を含むことができる。いくつかの実施形態では、発話を、複数の分節に分節化することができ、これらの分節のうちのいくつかは音素に対応し、その他には音節に対応するものもある。いくつかの実施形態では、使用される分節化（例えば、音素ベース、音節ベース、またはその他）は、信頼度メトリックまたは精度メトリックに依拠することができる。発話は、発話の分節同士の間は無声音期間を含むこともできる。いくつかの例によれば、発話を視覚化する図像表現が生成され、この図像表現は、非専門家ユーザにとってより直感的となり得るように、またはよりユーザフレンドリーとなり得るように発話を視覚化したものであり、コンピューティング装置の画面上に表示される。発話を視覚化するために使用される図像表現は、1つまたは複数のオブジェクトを含むことができ、これらのオブジェクトの各々は、発話の1つの分節に対応する。図像表現を生成する際には、発話のそれぞれの分節の持続時間は、オブジェクトの長さによって表現され、発話のその分節の強さは、オブジェクトの幅によって表現される。発話の個々の分節を表現する個々のオブジェクトを、図像表現において互いに間隔を空けて配置することができ、その間隔は、対応する分節間の無声音期間に対応する。本明細書の実施形態では、それぞれのオブジェクトは、境界線を有し、2つの隣接するオブジェクトの境界線同士の間隔のサイズ（例えば、長さ）は、対応する分節同士の間隔の無声音期間の持続時間に対応する。いくつかの実施形態では、オブジェクトは、長方形、楕円形、たまご形、または他の規則的な幾何形状から選択される形状を有することができる。規則的な幾何形状は、1つまたは複数の軸を中心とした対称性を有する形状であってよい。いくつかの実施形態では、オブジェクトは、明確に画定され得る（例えば、境界線によって縁取りされ得る／輪郭が描かれ得る）限り、かつ対応する分節の持続時間および強さをそれぞれ表現するための長さおよび幅を有し得る限り、必ずしも規則的な幾何形状によって表現されていなくてもよい。

【 0 0 1 4 】

いくつかの実施形態では、発話を視覚化するために使用される図像表現は、基準フレームなどに対するオブジェクトの傾きまたは傾斜角によって分節のピッチ曲線を表現することをさらに含むことができ、なお、基準フレームは、表示してもよいが、多くの場合、表示しなくてもよい。本明細書の文脈では、ピッチ曲線は、ピッチパラメータとも称される、知覚される音声の高さまたはピッチに関連する1つまたは複数の物理的なパラメータの動きを表現することができる。ピッチ曲線の一例は、基本周波数の動きを表現する曲線であってよいが、本明細書の例は、このピッチパラメータのみに限定されているわけではな

10

20

30

40

50

い。いくつかの実施形態では、オブジェクトの傾きまたは傾斜角は、オブジェクトの長さ
に沿って変化することができ、それによって発話の所与の分節に関連するピッチ曲線の変
移を捕捉または反映することができる。さらなる実施形態では、ピッチパラメータのオフ
セット（例えば、基本周波数のオフセット）を、視覚化部において、基準フレームに対す
るオブジェクトの高さによって表現することができる。いくつかの実施形態では、分節に
対応する１つまたは複数の音の調音部位および／または調音方法に基づいてオブジェクト
の色を選択するなどにより、視覚化部を介して発話に関する追加的な情報を伝えることが
できる。例えば、複数の異なる音素にそれぞれ異なる色を割り当てることができる。いく
つかの実施形態では、オブジェクトの色を、分節における最初の音素に基づいて選択す
ることができる。いくつかの実施形態では、複数の異なる音素の音の調音部位および／ま
たは調音方法の共通性（例えば、２つの異なる音素の音を調音するために同じ調音器官を
使用していること）を、色の共通性（例えば、同じ色の異なる色調、および／またはそれ
以外では、１つの色グループにまとめることができる複数の色）によって反映することが
できる。発話の直感的かつユーザフレンドリーな視覚化部を提供するために、種々異なる他
の組み合わせおよびバリエーションを使用することができる。発話のコンピュータ生成に
よる視覚化部を提供するための本明細書で説明されている方法は、例えば、コンピュー
ティング装置によって実行された場合に、本明細書の任意の例に従って発話の図像表現を
生成および／または表示することをコンピューティング装置に実施させるための命令の形態
で、コンピュータ可読媒体において具現化可能である。

【 0 0 1 5 】

図 1 は、本開示の実施形態による装置 1 0 の簡略化されたブロック図である。装置 1 0
は、部分的にスマートフォン、携帯型コンピューティング装置、ラップトップコンピュー
タ、ゲームコンソール、またはデスクトップコンピュータによって実装可能である。装置
1 0 を、任意の他の適切なコンピューティング装置によって実装してもよい。いくつかの
実施形態では、装置 1 0 は、プロセッサ 1 1 と、プロセッサ 1 1 に結合されたメモリ 1 2
と、同じくプロセッサ 1 1 に結合された、いくつかの例ではタッチ画面であってよいディ
スプレイ画面 1 3 とを含む。装置は、１つまたは複数の入力装置 1 6 と、外部通信インタ
ーフェース（例えば、無線送信機／受信機（ $T \times R$ ）1 7）と、１つまたは複数の出力
装置 1 9（例えば、ディスプレイ画面 1 3 および音響出力部 1 5）とをさらに含むこと
ができる。本願では、システムのコンポーネント（例えば、プロセッサ 1 1 およびメモリ
1 2 のような装置 1 0 のコンポーネント）を説明する際に単数形 “a” または “an” を参照して
いるが、これらのコンポーネント（例えば、プロセッサおよび／またはメモリ）のいづれ
も、本明細書で説明されているコンポーネントの機能性を提供するために（例えば、並列
にまたは他の適切な配置で）動作可能に配置されている１つまたは複数の個々のそのよう
なコンポーネントを含んでもよいことは理解されるであろう。例えば、メモリの場合には
、例えば並列に配置されていて、同じまたは異なる種類のデータを同じまたは異なる保存
時間で保存することができる複数のメモリ装置によって、メモリ 1 2 を実装することが
できる。いくつかの例では、ディスプレイ画面 1 3 を、（例えば、ディスプレイ画面 1 3 上
のグラフィックスおよびビデオデータの表示を制御するために）ディスプレイ画面 1 3 の
表示動作を制御するビデオプロセッサ（例えば、グラフィックスプロセッシングユニット
（GPU））に結合することができる。いくつかの実施形態では、ディスプレイ画面 1 3
は、タッチ画面であってよく、（例えば、ユーザ入力を介して受信した）ユーザインタ
ラクションデータをプロセッサ 1 1 に提供することができる。例えば、タッチ感応式のディ
スプレイ画面 1 3 は、タッチ画面の表面上の特定の領域のタップ、スワイプ等のようなユ
ーザのタッチ動作を検出することができる。タッチ画面は、検出されたタッチ動作に関
する情報をプロセッサ 1 1 に提供することができる。プロセッサ 1 1 は、場合によっては
タッチ動作に応答して、発話进行处理すること、および発話の視覚的表現を生成すること
を装置 1 0 に実施させることができる。したがって、装置 1 0 のタッチ感応式のディス
プレイ画面 1 3 は、入力装置 1 6 および出力装置 1 9 の両方として機能することができる。
いくつかの実施形態では、装置 1 0 は、１つまたは複数の追加的な入力装置 1 6（例えば、1

つまたは複数のボタン、キー、ポインティング装置等を含むことができる入力装置 18 および音響入力部 14) を含むことができる。いくつかの実施形態では、発話の処理は、部分的にプロセッサ 11 によって実施される。他の実施形態では、通信インターフェース (例えば、無線送信機 / 受信機 (Tx / Rx) 17) を介してプロセッサ 11 と通信している外部プロセッサによって、発話を処理することができる。無線送信機 / 受信機 (Tx / Rx) 17 は、モバイルネットワーク (例えば、3G、4G、5G、LTE、Wi-Fi 等) を使用して装置 10 とインターネットとの通信を容易にすることができるか、またはピア・ツー・ピア接続を使用して装置 10 と他の装置との通信を容易にすることができる。

【0016】

図示のように、装置 10 は、音響入力部 14 および音響出力部 15 を含むことができる。本願は、「1つの(an)」音響入力部および「1つの(an)」音響出力部に言及しているが、これらのコンポーネント (例えば、マイクロフォン入力部、音響出力部) のいずれも 1 つまたは複数を含んでもよいことが理解されるであろう。例えば、装置 10 は、内部および / または外部マイクロフォン用の 1 つまたは複数の音響入力部、内部および / または外部スピーカー用および / またはフォーンジャック用の 1 つまたは複数の音響出力部を含むことができる。いくつかの例では、音響入力部 14 および音響出力部 15 を、音響入力部 14 からの音響入力信号または音響出力部 15 への音響出力信号の音響信号処理を制御する 1 つまたは複数の音響信号プロセッサに結合することができる。したがって、音響入力部 14 および音響出力部 15 を、音響 DSP を介してプロセッサ 11 に動作可能に結合することができる。プロセッサ 11 は、音響入力信号から変換された音響データを録音すること、または音響出力信号を提供することによって音響データを再生することを装置 10 に実施させることができる。

【0017】

図 2A は、装置 10 によって (例えば、少なくとも部分的にプロセッサ 11 によって) 実施することができる、本開示のいくつかの実施形態による発話を視覚化するためのプロセス 200 のフロー図である。装置 10 は、ステップ S20 において発話入力を受信することができる。発話入力は、ユーザによる単語、フレーズ、またはその他の発語または発声であってよい。発話入力は、事前に録音および保存された発声 (例えば、基準発話) であってよい。発話入力は、装置 10 によって音響信号 (すなわち、発語または発声を表現する波形信号 (または単に波形)) として受信可能である。ブロック S21 に示されているように、公知のまたは後々開発される任意の発話認識技術を実装することができる発話エンジンが、発話入力 (すなわち、音響信号) を処理し、発話を分節化して、テキスト表現を取得することができる。追加的または代替的に、発話エンジンは、発話入力のスペクトログラムを出力してもよい。他の例では、スペクトログラムを、発話認識とは独立して、ここでも現在公知のまたは後々開発される技術を使用して取得してもよい。いくつかの実施形態では、発話入力のスペクトログラム表現を生成または取得することができるが、発話入力のスペクトログラム表現は、本明細書の視覚化エンジンの動作のために必須のものではない。いくつかの実施形態では、代替的または追加的に、ブロック S21 で発声に対して実施される任意の発話認識とは独立して、発声と共に基準テキストを提供してもよい。

【0018】

発話エンジンは、完全にまたは部分的に装置 10 のプロセッサ 11 によって実装可能である。いくつかの実施形態では、発話エンジンの少なくとも一部を、装置 10 からリモートに位置する、装置 10 に通信可能に結合されているプロセッサによって、例えば、装置 10 と無線通信しているサーバのプロセッサによって実装することができる。発話エンジンを、プログラム (例えば、コンピュータ可読媒体上に保存された命令) として実装することができ、このプログラムを、装置 10 にローカルに保存して実行してもよいし、リモートに保存して装置 10 によってローカルに実行してもよいし、またはこのプログラムの少なくとも一部をリモートのコンピューティング装置 (例えば、サーバ) に保存して実行してもよい。装置 10 はさらに、発話視覚化エンジン (speech visualization engine :

10

20

30

40

50

S V E) を実装することができ、この発話視覚化エンジン (S V E) も同様に、ローカルまたはリモートに (例えば、サーバ上、クラウド上に) 保存して少なくとも部分的に装置 1 0 によってローカルに実行することができるプログラムとして実装することができる。例えば、S V E は、プロセッサ 1 1 によってローカルに実行され、実行された場合に、本明細書の任意の例による視覚化プロセスを実施することができる。いくつかの例では、発話認識プロセスの一部であってよい発話の分節化 (S 2 2) を、ローカルに (例えば、プロセッサ 1 1 によって) 実施してもよいし、またはリモートに (例えば、リモート / クラウドサーバのプロセッサによって) 実施してもよい。分節化された発話入力 of 視覚的表出を生成するための視覚化プロセスは、プロセッサ 1 1 によってローカルに実施可能である。いくつかの例では、S V E のコンポーネントを、装置 1 0 に通信可能に結合された外部メモリ記憶装置 (例えば、U S B キーメモリ、クラウドに常駐するサーバのメモリ装置) にプログラムコードとして保存することができる。プロセス 2 0 0 のいずれかの部分 (例えば、分節化部分) がリモートに (例えば、クラウドで) 実行される場合には、視覚的表出を生成するための情報 (例えば、分節の特性、ピッチ情報等) を、装置の外部通信インターフェースを介して (例えば、無線送信機 / 受信機 1 7 または有線接続を介して) 装置に通信することができる。

10

【 0 0 1 9 】

(例えば、プロセッサ 1 1 によって発話入力として受信された) 発声を視覚的に表出するために、発話入力 that 分節化される。発話入力を分節に分解することを含む分節化は、装置 1 0 のプロセッサ 1 1 または別のプロセッサによって実行することができる発話エンジンによって実施可能である。例えば、発話エンジンは、発話入力を分解して、この発話入力を音節単位に分節化することができる (ブロック S 2 2 を参照)。これは、音節レベルでの分節化と称されることがある。この段階では、それぞれの分節がテキスト表現における想定される音節に対応するように発話入力を分割することによって、音節単位への分節化を実施することができる。しかしながら、種々異なるユーザの発音、特に子音間に母音の挿入が生じることがある非ネイティブスピーカーの発音のばらつきに起因して、音節レベルで分節化された場合に単一の音節を含むことが予想される 1 つの分節化された単位が、実際には複数の音節を含んでいる可能性がある。なぜなら、発話のその分節は、何人かのユーザによって (例えば、母音が存在すべきではない場所に母音を挿入することにより) それぞれ異なるように発音されるからである。したがって、プロセス 2 0 0 は、ステップ S 2 3 で開始する精度チェックを含むことができる。音節レベルでの分節化が完了すると (S 2 2)、プロセス 2 0 0 は、分節化された音節単位に含まれている音素が、その音節の予想される音素と実質的に一致するかどうかを判定するなどにより、音節レベルでの分節化の精度を判定することができる。例えば、プロセス 2 0 0 は、関連する音素を含んでいる音節単位または分節を、音素の基準配列と比較することができる。音素の基準配列は、テキスト表現に基づいて、一般的に使用されている辞書に列挙されている国際音声記号 (I P A) を使用するか、ネイティブスピーカーによる基準発話の録音を手動で注釈するか、またはネイティブスピーカーによる基準発話の録音に対して発話認識を実行することによって取得可能である。いくつかの実施形態では、基準発話の発音をより正確に表現するため (例えば、音の縮約を表現するため)、かつ / または I P A 記号によって提供される以上の追加的なガイダンスをユーザに提供するために、I P A 記号の 1 つまたは複数の修正版を使用することができる。例えば、I P A 記号をさらに注釈するためのマークまたは他のメカニズムを使用してもよい。いくつかの実施形態では、I P A 記号の修正版は、記号を太字で表現すること、より小さい文字対より大きい文字で表現すること等を含むことができる。音節単位または分節における音素が、音素の基準配列に非常に対応していると判定された場合 (イエス: S 2 3) には、プロセス 2 0 0 は、音節分節化が十分な精度であると判定し、(S 2 4 における) 音節分節の図像表現 (視覚的表出とも称される) の生成に関連するステップに進む。音節分節が音素の基準配列にさほど対応していないと判定されるなどにより、音節分節化の精度が低い場合 (ノー: S 2 3) には、音素レベルでの分節化を継続することができる (S 2 5)。ここでは、ステップ S 2 2 における音節

20

30

40

50

レベルでの分節化からの想定された音節単位または分節（例えば、基準配列との対応性が低い単位）が音素レベルで再検討され、単一の音節に対応することが想定された音節分節が、1つの分節内に複数の母音が識別されるなどによって実際には2つ以上の音節を含んでいると判定された場合には（イエス：S 2 6）、それぞれの分節が1つの母音を含むように、この分節を2つの分節に分割することができる（S 2 7）。それぞれの分節が1つの音節を含むことが保証された後、装置（例えば、プロセッサ 1 1）は、音節 / 音素分節に基づいて、発話入力 of 視覚的表出を生成することができる（S 2 4）。視覚的表出を表示する際には、完全な視覚化部（例えば、発話入力のために生成された全てのオブジェクト）を一度に表示してもよいし、またはオブジェクトの表示をアニメーションの形態で（例えば、先行するオブジェクトが表示された後に連続するオブジェクトが順次に表示される）実現してもよい。

10

【0 0 2 0】

図 2 B は、本開示のいくつかの実施形態による、発話の分節の視覚的表出または視覚的表現を生成するためのプロセス 2 4 0 のフロー図である。プロセス 2 4 0 は、少なくとも部分的に図 2 A のプロセスのステップ S 2 4 を実装するために使用可能である。プロセス 2 4 0 を、図 2 A のプロセスを介して抽出された分節に対して実施してもよいし、または従来技術のような別の異なるプロセスによって抽出された分節に対して実施してもよい。プロセス 2 4 0 は、例えば装置 1 0 のプロセッサ 1 1 によってローカルに実行される、本開示による S V E によって実施可能である。図 2 B のプロセスを使用して、発話入力におけるそれぞれの音節分節ごとに1つの図像オブジェクトが作成されるように、発話の視覚的表現を生成することができる（ブロック S 2 4 1 を参照）。ステップ S 2 4 1 は、分節のための規則的な形状のオブジェクト（例えば、楕円形、長方形、たまご形、またはその他）のような任意の適切な形状のオブジェクトからオブジェクトを選択することと、図像オブジェクトの各々の長さ、幅、およびオブションとして傾斜角、垂直方向の位置、色等のようなパラメータを設定することを含むことができる。このステップ S 2 4 1 は、それぞれの分節（例えば、音節または音素のようなそれぞれの分節化された有声音単位）が1つのオブジェクトによって視覚的に表現されるように、発話入力のそれぞれの分節ごとに実施可能である。好ましくは、見映えを良くするために、所与の視覚化された発話入力における全ての分節に対して同じ形状のオブジェクト（例えば、全てたまご形、または全て長方形）を使用することができる。しかしながら、任意の所与の視覚化部（例えば、所与のフレーズを視覚化する場合）または一連の視覚化部に対してそれぞれ異なる形状のオブジェクトを使用してもよいことが企図されている。いくつかの実施形態では、視覚化部に対して使用されるオブジェクトの種類（例えば、長方形、たまご形等）を、ユーザによって構成可能とすることができる。他の例では、視覚化部に対して使用されるオブジェクトの種類を、S V E に事前にプログラミングしておくことができる。

20

30

【0 0 2 1】

図 2 B を再び参照すると共に、例示的な視覚化部 2 0 4 を示す図 2 C も参照すると、任意の所与のオブジェクト 2 0 1 の長さ（L）を、所与の分節の持続時間を表現するように、または所与の分節の持続時間に対応するように設定することができる。これにより、発話入力の分節の各々の持続時間が取得される（ステップ S 2 4 1 1 において）。例えば、発話入力に対応する波形および / またはスペクトログラムから、開始時間および終了時間、ひいては発話入力のいずれかの音節 / 音素分節の持続時間を取得することができる。（例えば、波形および / またはスペクトログラムから、場合によっては発話認識プロセスの最中に）音節 / 音素分節の各々の強さを取得することもでき、それぞれの図像オブジェクトの幅（W）を、それぞれの分節の強さに従って設定することができる（S 2 4 1 2 において）。ステップ S 2 4 1 1 および S 2 4 1 2 は、任意の順序で実行してよい。この基本的な韻律情報がそれぞれのオブジェクトに取り込まれた状態で、オブジェクトの図像表現をディスプレイ画面上に表示することによって発話入力の視覚化部 2 0 4 を生成および表示することができる（S 2 4 2）。いくつかの実施形態では、プロセスは、発話入力の視覚的表現をさらに調整するための追加的なオブションのステップ（S 2 4 3）を含むことが

40

50

できる。さらに説明されるように、発話入力に関する追加的な韻律情報を伝達するために、図像オブジェクトの他の態様と、図像オブジェクトの相対的な配置とを任意に調整することができる。例えば、オブジェクト同士を、分節同士の間の無声音期間（例えば、検出可能な音節または音素に対応すると判定されなかった期間）に基づいて互いに間隔を空けて配置することができる。いくつかの例では、オブジェクトの傾きまたは傾斜角を、発話入力のピッチ曲線を反映するように設定することができる。さらに別の例では、個々の図像オブジェクトを、垂直方向に整列させなくてもよく、しかも、所与の分節の基本周波数のピッチ高さまたはオフセットのような追加的な韻律情報を伝達するために（例えば、互いに対してかつ／または基準フレームに対して）オフセットさせることができる。さらに別の例では、オブジェクトの色を、分節に関連する音の調音部位および／または調音方法

10

【 0 0 2 2 】

図 2 B および図 2 C に戻ると、発話入力の図像表現（または視覚化部）2 0 4 が画面（例えば、装置 1 0 のディスプレイ画面 1 3）上に表示され（S 2 4 2 において）、この図像表現 2 0 4 は、発話入力の分節の各々を表現する複数の図像オブジェクトを含む。いくつかの実施形態では、視覚化部 2 0 4 は、所与の発話入力の全ての分節が分析されて、対応するオブジェクト 2 0 1 が作成された後に表示される。他の実施形態では、所与の発声（例えば、発話されたフレーズ）の視覚化部 2 0 4 を構築するために発話入力を処理しながら、図像表現（例えば、個々のオブジェクト 2 0 1）を順次に表示することができる。すなわち、1 つまたは複数の図像オブジェクト 2 0 1 を、関連する分節が処理されて、オブジェクトのパラメータ（例えば、長さ、幅、色、傾き、垂直方向の位置、間隔等）が決定されるとすぐに表示することができる。図 2 C は、本開示による発話の視覚的表現 2 0 4（視覚的表出または視覚化部 2 0 4 とも称される）の一例を示す。図 2 C の例では、発話入力におけるそれぞれの識別された分節に対応するそれぞれの図像オブジェクト 2 0 1 は、規則的な幾何形状、この場合には楕円形を有する 2 次元のオブジェクト 2 0 1 である。図像オブジェクト 2 0 1 は、それぞれ境界線によって画定されており、この例では、画面上の時間軸および周波数軸によって画定された基準フレームに対して示されている。図 2 C には、本例の理解を容易にするために基準フレームの軸が示されているが、視覚化部 2 0 4 が（例えば、装置 1 0 のディスプレイ画面 1 3 上で）ユーザに提供される際に、基準フレームを表示しなくてもよいことは理解されるであろう。図像オブジェクトは、任意の適切な形状を有することができる。例えば、直感的で見やすい視覚化のために、図像オブジェクトの形状を、長方形、楕円形、たまご形、または任意の他の規則的な幾何形状から選択することができる。少なくとも 1 つの対称線を備えた実質的にあらゆる幾何形状（例えば、涙滴形、台形、またはその他）を使用することができる。いくつかの実施形態では、所与のオブジェクト 2 0 1 の長手方向（ひいては長さ）は、本例のように実質的に一直線上に存在することができる。しかし、他の例では、長手方向がカーブに沿っていてもよく、したがって、オブジェクトの傾斜角または傾きがオブジェクトの長さに沿って変化してもよい。このことは、単一の分節内におけるピッチの変動を表現するために使用可能である。視覚化部 2 0 4 の連続するオブジェクトは、発声の全ての分節が画面上に視覚的に表現されるように、発話入力の連続する分節に関連付けられる。本例のようないくつかの実施形態では、オブジェクト同士を、発話入力の無声音期間に対応する距離の分だけ間隔を空けて配置することができる。例えば、図 2 C では、複数の図像オブジェクトは、これらの図像オブジェクトの各々の開始端部を、時間軸に沿ってオフセットされた位置に整列させることによって画面上に水平に配置されており、なお、このオフセットは、それぞれの分節の開始時間に基づいている。上述したようにオブジェクト同士を、所定の間隔の分だけ離間させることができ、この間隔は、分節の明瞭な視覚的表現を提供することができる。かつ／または追加的な韻律情報（例えば、有声音期間と有声音期間の間の休止の持続時間）を伝達することができる。言い換えれば、2 つの隣接するオブジェクトの境界線を、いくつかの例では、これら 2 つの隣接するオブジェクトに関連する 2 つの分節の間の無声音期間の持続時間に基づいた距離の分だけ間隔を空けて配置することができる。図 2 C

20

30

40

50

の場合には、ネイティブスピーカーによって発声された “What if something goes wrong” というフレーズの発話入力の視覚化例が示されており、この発話入力は、図 2 C の例では、それぞれ I P A 文字列における

【数 1】

[wʌt] [ɪf] [sʌm] [θɪŋ] [ɡoʊz] [rɒŋ]

として注釈および表現される分節 # 1 ~ 6 を含んでいることが特定されている。図 2 C の視覚化例において見て取れるように、最後の分節 “wrong” は、図 2 C のオブジェクト 2 0 1 - 6 の長さによって反映されているとおり、ネイティブスピーカーによって発声された場合には、典型的には最も長い時間を要する。いくつかの実施形態では、それぞれの分節、すなわち音節または音素分節に対応するそれぞれのオブジェクトを、追加的に、各自の対応する I P A 注釈または I P A 記号と共に表示することができる。いくつかの実施形態では、I P A 注釈または I P A 記号を、学習者によって容易に認識される種々異なるフォントサイズを使用して、種々異なるフォントスタイルを使用して、太字、斜体、下線のような種々異なる種類の強調表示を使用して、またはアクセント等を表現する追加的なマークを使用して表現することができる。

【0 0 2 3】

ここで図 2 D ~ 図 2 G も参照すると、発話の視覚的表出または視覚的表現の種々異なるバリエーションが示されている。図 2 D ~ 図 2 G の視覚的表現の各々は、同じ発話入力（例えば、“What if something goes wrong” というフレーズの同じ発声）を視覚化したものである。前述したように、図像オブジェクト 2 0 1 の種々異なる態様と、互いに対するかつ／または基準フレーム（図示せず）に対する図像オブジェクト 2 0 1 の相対的な整列とは、視覚化部の直感的でユーザフレンドリーな性質をなおも維持しながら、種々異なるレベルの豊かさを有する（例えば、種々異なる量または種類の韻律情報を伝達する）発話の視覚化部を提供するために変更可能である。図 2 D では、発話入力の視覚的表出（または視覚化部）2 0 4 - 1 は、それぞれのオブジェクト 2 0 1 の長さ（L）および幅（W）を介してそれぞれの分節（例えば、前述したように分節化されたそれぞれの音節または音素単位）の持続時間および強さを伝えるだけでなく、オブジェクトの傾斜または傾きを変化させることによってピッチ情報と、オブジェクト同士の間隔を介して発声休止情報と、それぞれのオブジェクトの色を適切に選択することによって音素情報とを伝える。図 2 E には、同じ発話入力の簡略化された表現 2 0 4 - 2 が示されており、ここでは、持続時間および強さのような特定の分節情報が、それぞれのオブジェクトのサイズを介して伝えられ、休止情報および音素情報が、オブジェクトの間隔および色を介して伝えられる。図 2 E の例には、ピッチ曲線情報が含まれていないが、図 2 E に類似した他の例では、いくつかのピッチ曲線情報（例えば、基本周波数）を伝達するために、図 2 D のようにオブジェクトの垂直方向のオフセットを変化させることなく、オブジェクトの傾きをさらに変化させることができ、それにより、いくつかの他のピッチ曲線情報（例えば、分節の基本周波数のオフセット）が省略される。図 2 F は、発話入力の視覚的表現の別の例 2 0 4 - 3 を示し、この例は、図 2 C の例に類似しているが、図 2 C で使用された楕円形とは異なる形状のたまご形を利用している。図 2 F では、持続時間および強さのような基本的な分節情報が、それぞれのオブジェクトのサイズを介して伝えられ、無声音期間（例えば、発話の発声における休止）の持続時間が、オブジェクト同士の間隔を介して伝えられる。視覚化部からピッチ曲線情報を省略してもよいし、または少なくともいくつかのピッチ曲線情報を省略してもよい。上述したように、ピッチに関する少なくともいくつかの情報を伝えるために、図 2 F のオブジェクトの傾きを、これらのオブジェクトの垂直方向のオフセットを変化させることなく変化させることができる。所与の視覚化部の全てのオブジェクトを、同じ色で表示することができ、ここではグレースケール色（例えば、黒色）が示されているが、単色の視覚化部が、任意の色（例えば、任意の R G B 色または C M Y K 色）を利用してもよいことが理解されるであろう。図 2 G に示されているよう

なさらに別のバリエーションでは、図像オブジェクトを、（例えば、持続時間および強さ情報を伝達するために）種々のサイズで、（例えば、音素情報を伝達するために）種々の色で表示することができるが、ピッチ情報および休止情報を省略することができる。図 2 G に示されているように、ここでは、オブジェクト同士が実質的に互いに隣接するように配置されている（例えば、隣接するオブジェクトの境界線は、たとえ無声音期間が存在していても、隣接する分節（例えば、音節単位）同士の間は無声音期間の持続時間に関係なく、互いに隣接または接触することができる）。理解されるように、少なくともいくつかの韻律情報を伝達する発話の、簡略化されたユーザフレンドリーな視覚化部を提供するために、本明細書で説明されている視覚化技術の特徴を組み合わせた他のバリエーションを使用することができる。

10

【 0 0 2 4 】

上述したようにオプションとして、発話の視覚的表現のそれぞれの図像オブジェクトに 1 つの色を割り当てることができ、いくつかの実施形態では、色の割り当ては、その分節に関連する音の調音部位および / または調音方法に基づくことができる。例えば、色は、所与の分節によって表現される特定の音節または音素に基づくことができる。分節（例えば、音節単位）が複数の音素を有している例では、オブジェクトの色を、その分節の最初の音素に基づいて選択することができる。いくつかの実施形態では、調音部位および / または調音方法の共通性を、オブジェクトために使用される色の共通性によって反映することができる。例えば、1 つの共通する調音部位を有する音（例えば、両唇音、唇歯音等）を備えた分節に、それぞれ同じ色グループの色（例えば、図 3 B に示されているように、種々異なる色調またはニュアンスのピンクまたはバイオレット、または種々異なる色調のオレンジ）を割り当てることができる。

20

【 0 0 2 5 】

図 3 A は、発話 3 0 4 の視覚的表現のオブジェクト 3 0 1 に色を割り当てることを含む、本開示による分節の視覚的表現を生成するためのプロセス 3 0 0 のフロー図を示す。オブジェクトに色を割り当てるプロセス 3 0 0 は、それぞれの分節ごとにオブジェクトを作成するプロセスにおける（例えば、プロセス 2 4 0 のステップ S 2 4 1 における）追加的なオプションのプロセス / ステップとして含まれていてもよい。ステップ S 3 0 に示されているように、S V E（例えば、プロセッサ 1 1）は、そのオブジェクトに関連する分節の音素に基づいてオブジェクトに色を割り当てることことができる。1 つの分節が複数の音素を含んでいる場合には、関連する分節の最初の音素に基づいてオブジェクトに色を割り当てることことができる（S 3 2）。そのために、S V E（例えば、プロセッサ 1 1）は、それぞれの分節における最初の音素を特定することことができる（S 3 1）。音素の実際の検出は、分節化プロセスにおいて実施可能である。代替的に、分節が複数の音素を有するかどうかを識別するため、かつ / または分節における最初の音素を識別するために、それぞれの音節分節ごとに音素分節化を実施することことができる。S V E（例えば、プロセッサ 1 1）は、オブジェクトに割り当てるべき色を選択する際に、ルックアップテーブルを参照することことができる。いくつかの実施形態では、ルックアップテーブルは、音素または分節の最初の音素が識別されるとオブジェクトに適切な色を割り当てることできるように、それぞれの音素ごとに一意の色を指定することことができる。この例では、オブジェクトの色を選択するために音素が使用されているが、他の例では、調音部位および / または調音方法に結びついた別の異なるパラメータを色の選択のために使用することことができる。例えば、それぞれの音素に一意の色を割り当てる代わりに、同じ調音部位（例えば、唇音、唇歯音など）に関連する全ての音に 1 つの同じ色を割り当ててもよい。したがって、そのような例では、ルックアップテーブルは、代替的にまたは追加的に、音の種々異なる調音部位および / または調音方法に対して 1 つの対応する色を識別することことができる。

30

40

【 0 0 2 6 】

図 3 B には、このようなカラーテーブルの例が、少なくとも部分的に視覚的に表現されている。図 3 B の図は、本開示の実施形態による、色と、子音を含む音素と、子音に関連する声道内の調音部位（位置）との間の関係を示す。色のグラデーションを、関連する子

50

音に関連付けて割り当てることができ、例えば、その関係は、声道内の調音部位および調音方法に基づいている。例えば、唇で作られる唇音 [p] [b] [m] および [w] を、同じグループにグループ化して、同じ色グループ（例えば、ピンク - 紫の色グループ）に関連付けることができ、これらの音素の各々は、無声破裂音、有声破裂音、鼻音、および接近音のように調音方法が異なっているので、これらの音素の各々を、この色グループにおけるそれぞれ異なる色調またはグラデーションに関連付けることができ、すなわち、本例では、これらの唇音に、ピンク - 紫の種々異なるグラデーションの色形を割り当てることができる。同様に、対応する母音に割り当てられた色の段階的なシフトを設けることができ、この段階的なシフトは、典型的には比較的低いフォルマント周波数（例えば、F₁ および F₂）として抽出される母音にとって特有の共鳴に対して影響を与える、話者の声道の位置および開きの段階的なシフトに基づくことができる。特定の色および関連付けは、単なる一例として提供されているに過ぎず、他の実施形態では、色と音素 / 音との間の別の異なる関連付けを使用してもよいことが理解されるであろう。それぞれのオブジェクトに色が割り当てられた後（S 3 2）、発話の豊かな視覚的表出を提供するために、視覚化部 3 0 4 のオブジェクトを適切な色で表示することができる。

【 0 0 2 7 】

図 3 C は、本開示の実施形態による、発話の生成された視覚的表現 3 0 4 のタイミング図である。図 3 C の視覚的表出 3 0 4 は、図 2 C に示されている “ What if something goes wrong ” というフレーズの同じ発声であり、したがって、図像オブジェクト 3 0 1 のサイズおよび配置は、図 2 C のオブジェクト 2 0 1 のサイズおよび配置と同じであり、こ

【 数 2 】

[w] [l] [s] [θ] [g] [r]

であり、したがって、分節 # 1 ~ 6 に関連するオブジェクトは、図 3 B に示されている音素 - 色の関連付けに従ってそれぞれ紫、黄、青、黄緑、暗灰、濃紺の色によって色分けされている。オブションとして、視覚化部（例えば、3 0 4、2 0 4 - 1、2 0 4 - 4 等）によって提供される視覚的なガイダンスを、ユーザが読むことおよび慣れることを補助するための追加的なトレーニングリソースとして、図 3 B に示されている色の関連付けを、（例えば、ディスプレイ上で、または被印刷物の形態で）ユーザに提供することができる。

【 0 0 2 8 】

図 3 D は、本開示のさらなる実施形態による、発話の生成された視覚的表現 3 1 7 - 1 および 3 1 7 - 2 と、発話に関連する顔表現 3 1 8 - 1 および 3 1 8 - 2 とを含む画面 3 1 3 の概略図である。いくつかの実施形態では、画面 3 1 3 は、装置 1 0 のディスプレイ画面 1 3 であってよい。例えば、画面 3 1 3 は、タッチ画面であってよい。画面 3 1 3 は、ディスプレイウィンドウ 3 1 4 および 3 1 5 を表示することができる。ウィンドウ 3 1 4 は、本開示の実施形態による、発話の生成された視覚的表現 3 1 7 - 1 および 3 1 7 - 2 を表示することができる。いくつかの実施形態では、発話の生成された視覚的表現 3 1 7 - 1 および 3 1 7 - 2 は、発話の波形のようなタイミング図であってよい。いくつかの実施形態では、発話は、2 人の話者（例えば、図 3 D のチューターおよびユーザ 1）によって生成される同一のフレーズ（例えば、図 3 D の “ take care ”）の抜粋であってよい。いくつかの実施形態では、第 1 の生成された視覚的表現 3 1 7 - 1 は、言語のネイティブスピーカーまたは言語教師によって提供された基準発話を示すことができ、第 2 の生成された視覚的表現 3 1 7 - 2 は、ユーザの発話（例えば、学習者の発話）を示すことができる。いくつかの実施形態では、生成された視覚的表現 3 1 7 - 1 および 3 1 7 - 2 は、それぞれオブジェクト 3 1 9 - 1 1 および 3 1 9 - 1 2 ならびにオブジェクト 3 1 9 - 2 1 および 3 1 9 - 2 2 を含むことができる。これらのオブジェクトのうちの 1 つまたは複数に、場合によってはオブジェクト 3 1 9 - 1 1、3 1 9 - 1 2、3 1 9 - 2 1、および 3

19 - 22の各々に、色を割り当てることができる。発話におけるそれぞれ異なる音素（例えば、[t]または[k]）には、それぞれ異なる色（例えば、水色または灰色）を関連付けることができ、したがって、視覚的表現のそれぞれ異なるオブジェクトには、所与の発話の音素に対応するそれぞれ異なる色を割り当てることができる。画面313は、所与の発話において表現される1つまたは複数の音素の音の調音部位および/または調音方法に関するユーザガイダンスを提供する（例えば、アニメーションまたは静止画像の形態での）調音指示画像を提供するように構成可能である。例えば、画面313は、アイコン316を表示することができ、このアイコン316は、ユーザによって選択されると、例えば補助ウィンドウ315に調音指示画像を表示する。図3Dの2つのディスプレイウィンドウ314および315に示されているコンテンツ（例えば、視覚的表現317-1および317-2ならびに顔表現318-1および318-2）を、単一のウィンドウにおいて提示してもよいし、または本明細書の他の実施形態では、他の適切な数のディスプレイウィンドウにおいて提供してもよい。

10

【0029】

図3Dの具体的かつ非限定的な例を参照すると、システムは、調音指示が作動させられると、発話のそれぞれの音素ごとの、または音素の部分集合の（例えば、それぞれの音節の最初の音素の）それぞれの画像表現または顔表現318-1、318-2を表示することができる。それぞれの画像表現または顔表現318-1および318-2は、発話における1つまたは複数の音（例えば、図3Dの“take care”というフレーズまたは発話における[t]および[k]の音）の調音部位および/または調音方法を、オプションとして関連する波形と一緒に反映することができる。いくつかの実施形態では、調音指示は、基準発話を模倣するために発話をどのようにして適切に発音するべきかに関するガイダンスを提供するように、基準発話に適合させられている（例えば、基準発話の視覚化部要素を選択することによって呼び出されるか、または基準発話に近接して配置される）。調音指示（例えば、顔表現318-1および318-2）を、発話視覚化部の一部ではないアイコン316が選択されたことに応答して提示してもよいし、またはオブジェクト319-11および319-12のうちの1つまたは複数のような、発話視覚化部の要素が選択されることによって提示してもよい。いくつかの実施形態では、発話の視覚的表現317-1のオブジェクトのいずれかを選択すると、そのオブジェクトに関連する顔表現だけを表示させることができ、その一方で、アイコン316を選択すると、視覚的表現317-1のオブジェクトの各々に関連する顔表現を、例えば顔表現のシーケンスとして表示させることができる。所与のオブジェクトに関連する顔表現を、例えば、この所与のオブジェクトの色に対応する色を表示することによって、この所与のオブジェクトに視覚的に関連付けることができる。いくつかの実施形態では、顔表現318-1および318-2のうちの個々の顔表現は、静止していてもよいし、または所与の音を適切に発音するためにユーザがどのようにして唇、舌、口等を動かすべきかの手法のような、代表的な音の調音部位および/または調音方法を反映しているアニメーションまたは動画として表示されてもよい。

20

30

【0030】

発話入力のパッチ曲線を、本開示の原理に従って図像的に表現することができる。図4Aは、本開示のさらなる実施形態による、発話入力の見覚的表現404を生成するためのプロセス400のフロー図である。プロセス400は、図2Bのプロセス240の追加的なステップまたはプロセス（例えば、S243）を部分的に実装するために使用可能である。図4Aの例では、プロセス400は、発声のパッチ情報を伝達するようにオブジェクトを配置することを含み、したがって、発話入力のパッチ曲線を視覚的に表現するために使用可能である。他の例では、視覚化部（例えば、204、304等）を提供するためにプロセス240のステップS241で作成されるオブジェクトの相対的な配置は、種々異なる組み合わせ（例えば、プロセス400のステップまたは追加的なステップの組み合わせの構成要素）を含むことができる。プロセス400は、それぞれの分節ごとにピッチパラメータ（例えば、基本周波数、または聴者によるピッチの知覚を代表する他のパラメー

40

50

タ)を検出することを含むことができる(S 4 1)。従来の基本周波数のような、知覚される音声の高さに関連する1つまたは複数の物理的なパラメータ(ピッチパラメータ)の動きを表現するピッチ曲線を開発することができる。ピッチパラメータは、必ずしも基本周波数に限定されているわけではなく、聴者による発話の音声の高さの知覚に対して影響を与える可能性のある他の物理的または生理的なパラメータを、ピッチパラメータとして使用してもよい。検出されたピッチパラメータと、例えばピッチパラメータの増加または減少として検出されるピッチの上昇または下降の勾配のような、発話入力のパッチ曲線とに基づいて、それぞれのオブジェクトに傾き(または傾斜角)を割り当てることができる(S 4 2)。オブジェクトの傾きは、オブジェクトの長手方向と、基準水平軸(例えば、時間軸)との間の角度として見て取ることができる。いくつかの実施形態では、そこでプロセス400を終了することができ、その後、視覚化部のオブジェクト401を、各自のそれぞれの傾きと共に、ただし実質的に垂直方向に整列させられた状態で表示することができる。

10

【0031】

追加的にまたはオプションとして、プロセス400は、分節のピッチパラメータのオフセット(例えば、分節の基本周波数のオフセット)のような追加的なピッチ情報を伝達するために、(例えば、オブジェクト同士を、互いに対してかつ/または基準フレームに対して垂直方向にオフセットさせることによって)オブジェクトを垂直方向に配置することを含むことができる。このことを、ステップS 4 3およびS 4 4に示されているように、(例えば、互いに対するおよび/または基準フレームに対する)オブジェクトの相対的な垂直方向の位置によって視覚的に表現することができる。いくつかの例では、基準フレームであって、かつこの基準フレームに対して相対的に垂直方向のオフセットを決定することができるという基準フレームは、所定の基準線に基づくことができるか、または所与の発話入力に対して検出された最小のピッチパラメータに基づくことができる。図4Bは、図2Cおよび図3Cで視覚化されたものと同じ発話入力の波形405およびスペクトログラム407のタイミング図を示すが、ここでは、ピッチに関連する追加的な韻律情報を視覚化することが示されている。発話入力の生成された視覚的表現404は、スペクトログラム407に重畳された状態で示されている。観察され得るように、スペクトログラム407によって伝達される情報は、非専門家ユーザによって読み取ることが不可能ではないとしても困難である可能性があるが、その一方で、スペクトログラム407に含まれている韻律情報の少なくとも一部を伝達する視覚化部404は、非専門家ユーザによってより容易に理解することが可能である。本明細書では例示する目的でのみ示されている視覚化部404およびスペクトログラム407の重畳において、視覚化部404が発話入力の韻律に関する有用な情報をどのようにして非専門家ユーザに伝達することができるかを説明するために、オブジェクトは、青色の点の集合によって示されている実際の基本周波数曲線に視覚的に整列させられ、典型的には、熟練したユーザ/専門家ユーザによってスペクトログラムに抽出または追加することができる注釈に視覚的に整列される。

20

30

【0032】

図5Aおよび図5Bは、同じフレーズの第1および第2の発声の波形505aおよび505bならびにスペクトログラム507aおよび507bを示す。波形505aおよびスペクトログラム507aによって表現される第1の発声は、基準発声(例えば、言語学習アプリケーションの文脈での、例えばネイティブスピーカーによる発話入力)であってよい。波形505bおよびスペクトログラム507bによって表現される第2の発声は、ユーザ発声(例えば、言語学習の手本に続けて発話する学習者による発話入力)であってよい。図5Aおよび図5Bは、本開示に従って生成され、かつ対応するスペクトログラムに重畳された、第1および第2の発話入力の対応する視覚的表現504aおよび504bもそれぞれ示す。また、識別された有声音分節の各々の持続時間(例えば、分節持続時間506aおよび506b)と、分節の少なくとも一部の開始時間および/または終了時間とを含む、特定のタイミング情報も示されている。また、分節化の詳細(例えば、第1の発話入力の分節の記号表現509a、および第2の発話入力の分節の記号表現509b)も

40

50

示されている。図 5 A の第 1 の発話入力（例えば、ネイティブスピーカー）と比較すると、図 5 B の第 2 の発話入力（例えば、言語学習者）は、

【数 3】

[ɪf]

の代わりに [i] [h u] や、

【数 4】

[sʌm]

10

の代わりに [s a] [m u] のような、母音挿入によって作成された余分な音節分節を含む。これらの不一致は、オブジェクト同士の間に明瞭な間隔を有するオブジェクトの長さのような、発話の図像表現によって提供される時間情報によって良好に表現されており、したがって、非専門家ユーザによって容易に視認可能である。また、[f] の代わりに [h] や、[] の代わりに [s] のようないくつかの子音も、違ったように生成されている。これらの不一致も、音節分節を表現する色付きのオブジェクトによって良好に表現されており、したがって、非専門家ユーザによってその違いを容易に知覚することができる。また、オブジェクトの垂直方向の位置は、ピッチアクセントのタイミングの違いを示している（例えば、学習者の発声の場合には、10 番目の分節の比較的高くなっている垂直方向の位置によってピッチアクセントが見て取れるが、それに比べてネイティブスピーカーの発声の場合には、フレーズのその位置にピッチアクセントは存在しない）。上記の全ては、ユーザが自身の言語スキルを改善することを支援するために、基準発音と比較したときの自身の発音の違いをユーザが知覚することを補助するための直感的で理解しやすいツールを、本開示による発話の視覚化部によってどのようにして提供することができるかの例を提供するものである。

20

【0033】

図 6 A は、時間の関数としてプロットされた波形 605 およびスペクトログラム 607 を示し、このスペクトログラムには、本開示に従って生成された、ユーザ（例えば、言語学習者 - 学生 A）による発話入力の関連する視覚化部 604 - 1 が重畳されている。図 6 A の視覚化部 604 - 1 は、学習プロセス中の比較的初期の時間（例えば、第 1 日目）にユーザから取得された発話入力からのものであり、この視覚化部 604 - 1 は、図 6 B にも、例えば、本明細書の視覚化技術を実装する装置（例えば、装置 10）の画面上に表示され得るように（スペクトログラムから）分離された状態で示されている。図 6 C は、図 6 B と同じフレーズを発声する同じユーザ（例えば、言語学習者 - 学生 A）から、ただし学習プロセス中の比較的後期の時間（例えば、第 4 日目）において取得された、発話入力の視覚的表現 604 - 2 を示す。図 6 B の視覚的表現 604 - 1 と、図 6 C の視覚的表現 604 - 2 との視覚的な比較によって見て取れるように、両方の例において発話される単語が全く同じであるにもかかわらず、ユーザが同じフレーズをどのようにして発声するかの変化を、オブジェクトの図像表現の違いから容易に観察することができる。図 7 A は、図 6 A ~ 図 6 C と同じフレーズを発声するネイティブスピーカーによる発話入力の、時間の関数としてプロットされた波形 705 およびスペクトログラム 707 と、関連する視覚的表現 704 とを示し、この視覚的表現 704 は、図 7 A ではこの視覚的表現 704 の対応するスペクトログラムに重畳されている。図 7 B は、例えば、本明細書の視覚化技術を実装する装置（例えば、装置 10）の画面上に表示され得るように、図 7 A に示されているものと同じ視覚的表現を分離された状態で示す。ネイティブスピーカーによる発話入力の視覚的表現 704 と、ユーザ（例えば、言語学習者 - 学生 A）による発話入力の視覚的表現 604 - 1 および 604 - 2 との視覚的な比較から見て取れるように、2 人の話者の発声は、それぞれ異なる韻律を有している。したがって、ユーザは、自身の外国語の発声を改善するために（または自身の母国語の特定の方言またはアクセントのような発声を模

30

40

50

倣するために)、基準発話(例えば、図7Bに示されているようなネイティブスピーカーの発話)の視覚的表現704を参照または比較として使用することができる。図6Bにも示されているように、第1日目の発声(例えば、学生Aによる発話入力)の合計持続時間は、図6Cおよび図7Bの視覚化部604-2および704と比較すると、ユーザの初期時の十分に練習されていない発声における母音挿入(例えば、“f u”、“m + u”、“z u”、“u”、および“g + u”)に起因して著しく長くなっており、また、オブジェクトの個数の増加によって視認されるようにより多数の分節に分節化されている。また、図6Aおよび図6Bに表現されているオブジェクトの一部の色は、図6Cならびに図7Aおよび図7Bでは見受けられず、このことは、ユーザの発声(例えば、フレーズにおける音節に対応する調音方法および調音部位)が経時的に変化し、理想的には目標とする発声(例えば、ネイティブスピーカーの発声)により近似してきているということを実証している。他方で、図6Cの第4日目の学生Aによる発話の視覚的表現は、少なくともリズムに関しては、図7Bのネイティブスピーカーによる発話の視覚的表現により類似しているように見える。図6Aおよび図7Aの視覚化部を比較した場合の、オブジェクトの垂直方向の位置(または高さ)によって示されているピッチ曲線は、ネイティブスピーカーの発話入力のピッチ特性と比較したときの学習者の発話入力のピッチ特性の違いを実証している。図6Bの発声と図6Cの発声との間で見られるように母音挿入のいくつかは解消されているが、後々の時点でも(例えば、ある程度練習した後でも)、“ ”の代わりに“s”のようにいくつかの分節における子音の発音が、ネイティブスピーカーの基準発話とは依然として異なっているということが、視覚化部の比較から依然として明らかである。この視覚化技術を用いて、例えば、ユーザの発話視覚化部を基準発話の近傍(例えば、上または下)に表示することにより、ユーザ(例えば、言語学習者)は、自身の発話とネイティブスピーカーの発話との違いを容易に知覚することが可能となり、したがって、目標とする発声に向けて練習および改善することが可能となる。

【0034】

本明細書の例による言語学習アプリケーションまたは他の発話練習アプリケーションを実施する際のようないくつかの実施形態では、装置は、ユーザ(例えば、学習者)の視覚化部と、基準発話(例えば、ネイティブスピーカー)の視覚化部とを、これらの視覚化部の開始点(最初の端部)が実質的に垂直方向に整列させられた状態で表示することができる。図8A~図8Cは、本開示の実施形態による、発話の生成された視覚的表現804-1~804-3の概略図である。いくつかの実施形態では、基準発話の視覚化部(例えば、生成された視覚的表現804-1)を、ユーザの発話の視覚化部(例えば、生成された視覚的表現804-2または804-3)の近傍に(例えば、実質的に垂直方向に整列させられた状態で)表示することができる。図8A~図8Cの例では、生成された視覚的表現804-1~804-2は、3人の異なる話者(例えば、図8Aのチューター、図8Bのユーザ1、図8Cのユーザ2)によって生成された同じ発話、すなわち同一のフレーズ802またはその抜粋(例えば、図8Aの“No problem, I'll take care of him.”)の視覚化部を含む。いくつかの実施形態では、生成された視覚的表現804-1は、チューター(例えば、ネイティブスピーカーまたは言語教師)によって提供された基準発話における分節の視覚的表現であるオブジェクトを含むことができ、生成された視覚的表現804-2および804-3は、例えば、言語学習者(例えば、ユーザ1およびユーザ2)によって生成された発話における分節の視覚的表現であるオブジェクトを示すことができる。場合により、オブジェクトを、視覚化部が生成される元となった録音された発話のタイミング図および/または波形と一緒に(例えば、その上に重畳された状態で)表示することができる。いくつかの実施形態では、言語練習を容易にするために、ユーザ1に関連するコンピューティング装置の画面は、チューターの生成された視覚的表現804-1と、ユーザ1の生成された視覚的表現804-2とを、例えば実質的に垂直方向に整列させられた状態で表示することができる。他の実施形態では、2つの発話視覚化部を、横に隣り合って並べるなど、ディスプレイ上で近接するようにその他の手法で適切に配置してもよい。この例におけるユーザ1の視覚的表現804-2は、とりわけ、(例えば、チューター

の)基準発話の視覚的表現 804-1には存在しない可能性のある母音挿入(例えば、“b + u”、“v u”および“m + u”)に対応する可能性のあるオブジェクト 806-11, 806-12, および 806-13を含む。同様に、ユーザ2に関連するコンピューティング装置の画面は、チューターの生成された視覚的表現 804-1と、ユーザ2の生成された視覚的表現 804-3とをそれぞれ表示することができる。ユーザ2の視覚的表現 804-3は、とりわけ、基準発話の視覚的表現 804-1には存在しない可能性のある母音挿入(例えば、“b + u”、“m + u”、“v u”および“m + u”)に対応する可能性のあるオブジェクト 806-21, 806-22, 806-23, および 806-24を含む可能性がある。ユーザの視覚化された発話を基準発話の視覚化部に近接して提示することにより、システムは、ユーザ(例えば、学習者)が違いを識別して、単語、フレーズ等の「適切な」発音の模倣に向けた自身の進捗を把握することをさらに補助することができる。

10

【0035】

本明細書の例による言語学習アプリケーションまたは他の発話練習アプリケーションを実施する際のようないくつかの実施形態では、ユーザの発話の視覚化部を編集するために、装置を構成することができる。そのような編集は、ユーザの発話練習のための考えられる改善軌跡をユーザが視認することを補助するように、ユーザ入力(例えば、発声された発話に対してなされるべき編集をユーザが指定すること)に回答して実施されてもよいし、または装置によって自動的に実施されてもよい。本明細書で論じられるように、ユーザの発話の視覚化部と、基準発話の視覚化部とを同時に(例えば、画面上で垂直方向に、または横に隣り合って並べて)表示することができ、これにより、ユーザの発話の視覚化部と、基準発話(例えば、ネイティブスピーカー)の視覚化部との違いをユーザが復習することを可能にすることができる。その後、発話の選択された音節または他の分節の速度を変更(例えば、増加または低減)すること、音のレベルを低減または増幅すること、有声音分節と有声音分節との間の休止を短縮または延長すること、1つまたは複数の音を削除または低減する(例えば、日本語ネイティブスピーカーに典型的な母音挿入を除去すること、および/または他の修正を適用することなどによって、ユーザの発話の発声を編集することができる。図9は、本開示による、発話の視覚的表現を修正するフローの概略図である。図9は、(例えば、チューターの)基準発話の視覚的表現 902-1を示し、この視覚的表現 902-1を、ユーザの発話の1つまたは複数の視覚的表現(例えば、視覚的表現 902-2 ~ 902-4)と同時に表示することができ、これらの視覚的表現の各々は、発声された発話およびその分節の種々異なる特性を、オブジェクトを使用して視覚的に表現することができる。視覚的表現 902-1 ~ 902-4は、同じ発話のそれぞれ異なる発声、すなわち、複数の異なる話者(例えば、図9のチューターおよびユーザ)によって生成される、同じ単語またはフレーズのそれぞれ異なる発声に対応する。

20

30

【0036】

図9の例では、生成された視覚的表現 902-1は、図9ではチューターとしてラベル付けされている基準発話(例えば、ネイティブスピーカーまたは言語教師)の分節の視覚的表現である4つのオブジェクト 904-11 ~ 904-14を含む。ユーザによって発声された同じ発話の生成された視覚的表現 902-2は、8つのオブジェクト 904-21 ~ 904-28を含み、これらのオブジェクト 904-21 ~ 904-28は、同じ発話の発声の分節の視覚的表現であるが、ユーザ(例えば、言語学習者)によって生成されている。見て取れるように、ユーザの発声は、基準発話の発声には存在しない追加的なオブジェクトを含み、オブジェクトのうちの1つまたは複数のオブジェクトの特性(例えば、長さ、傾斜等)および/または間隔は、2つの視覚化部の間で異なっている。例えば、ユーザに関連する視覚化部のオブジェクト 904-21 ~ 904-24は、基準発話に含まれている音節を表現する基準発話のオブジェクト 904-11 ~ 904-14に対応する。他方で、ユーザの視覚化部のオブジェクト 904-25 ~ 904-28は、基準発話には存在せず、基準発話の一部ではない音節を表現している可能性がある。例えば、基準発話に含まれていない音節は、母音挿入または不正確な発音に起因する可能性がある。自身の発声におけるオブジェクトのうちの1つまたは複数に適用されるべき変更をユーザが

40

50

選択および指定する、またはユーザの発声と基準発声との間の違いをシステム（例えば、SVE）が自動的に決定するなどにより、ユーザの発声を編集し、この編集されたユーザ発声をフィードバックとして徐々に提示して、ユーザが自身の発声を徐々に改善することをアシストすることを、視覚的表現によって容易にすることができる。一例では、ユーザは、1つまたは複数の編集ステップを使用して、生成された視覚的表現 902-2 を編集することができる。例えば、第1の編集ステップにおいて、対応する音節の発音の速度を低減するために、オブジェクト 904-21, 904-23, および 904-24 を編集することができる。このことは、視覚的にはこれらのオブジェクトを拡大することに対応する。ユーザがオブジェクトをこのように直接的に編集したことに応答して、または先行するオブジェクト 904-21 が拡大された結果として、オブジェクト 904-25 が縮小される場合がある。したがって、編集後のユーザの発話の視覚化部 902-3 が再生される際には、オブジェクト 904-21, 904-23, 904-24, および 904-25 によって表現される音節は、それぞれより緩慢に、かつより高速に発音されることとなる。さらに、オブジェクト 904-23 と 904-24 との間にあるオブジェクト 904-26 および 904-27 と、最後のオブジェクト 904-28 とのような、基準発話には存在しない1つまたは複数のオブジェクトを削除または除去して、これによって編集済みのユーザの発声における音/音節の総数を低減するなどの、さらなる編集を行うことができる。同じ発話のユーザによる修正された発声を表現する視覚的表現（例えば、902-3 および 902-4）を、表示するために生成することができる。編集プロセスを、（例えば、「ユーザオリジナル」の発声から「2回目の編集後のユーザ」の発声に到達するまでの）1回のステップで実施してもよいし、または図示の例に示されているように複数のステップで実施してもよく、これにより、ユーザが練習を継続する際に目標とする徐々の改善のためのガイダンスを提供することができる。図9の例では、2回目の編集ステップが示されており、ここでは、1回目の編集済みのユーザ発声からのオブジェクト 904-35 が除去され、基準発話の場合と同数のオブジェクト（904-41~904-44）を含んでいる、視覚的表現 902-4 によって示されている発声に到達するために、オブジェクト 904-33 および/または 904-34 の速度をさらに調節（例えば、増加）することができる。したがって、基準発話に含まれているオブジェクト 904-11~904-14 に対応するオブジェクト 904-31~904-34 を含んでいる最終的な編集済みの発話の発声は、たとえ別の異なるユーザによるものであっても、基準発話に取り込まれたものと実質的に同様に発音される同数の音節を含むことができる。

【0037】

図9の例は、速度の変更、音節の削除/削減、音節の開始または終了のタイミングの変更を含む修正を例示しているが、本開示によるシステムによって提供される修正は、本明細書に具体的に例示されたものに限定されていない可能性がある。例えば、装置は、種々異なる他の修正またはそれらの任意の適切な組み合わせを可能にすることができ、例えば、それぞれの音のレベルを低減または増幅すること、音と音の間の休止を短縮または延長すること等を可能にすることができる。

【0038】

本発明の実施形態は、言語学習システムまたは言語学習アプリケーションを提供する装置（例えば、コンピューティング装置）によって実装可能である。図10A~10Dを参照しながら、例示的な実施形態がさらに説明されており、図10A~10Dは、本開示による、発話の視覚的表現を生成および/または提供するように構成されたコンピューティング装置のディスプレイ画面の画面キャプチャを示す。コンピューティング装置は、（タブレットまたはスマートフォンのような）携帯型コンピューティング装置であってよく、タッチ画面を含むことができる。本明細書の任意の例による発話の視覚的表現は、コンピューティング装置のタッチ画面上に表示可能である。例えば、図10A~10Dに示されているユーザインターフェースの画面ショットを、図1の装置10のタッチ画面上に表示することができる。他の実施形態では、視覚化部を、非タッチ感応式のディスプレイ画面上に提供し、タッチ画面とは異なる入力装置を介してユーザ入力を受信してもよい。装置

10

20

30

40

50

は、言語学習システムのプログラムを実行することができ、このプログラムの1つのコンポーネントは、発話の視覚的表現を生成することであってよい。種々異なる種類の発話を、言語学習プログラムの一部として視覚化することができる。例えば、図10Aに示されているように、装置（例えば、スマートフォン）のプロセッサは、（例えば、アプリケーション（「アプリ」）としてメモリ12に保存されている）コンピュータ可読命令の形態で具現化することができる、本明細書で説明されている視覚化プロセスを使用して、基準発話の簡略化された視覚化部1004aを生成することができ、この簡略化された視覚化部1004aも、メモリ12に保存することができる。図10Aに示されている画面ショット1002-1では、装置は、基準発話、例えばネイティブスピーカーによって提供された発話の視覚化部をタッチ画面上に表示している。視覚化部1004aが表示される前に、視覚化部1004aと一緒に、または視覚化部1004aが表示された後に、簡略化された視覚化部1004aに加えて基準発話の音響表現（例えば、音響再生）も、オプションとしてユーザに提供することができる。ユーザ命令に応答して（例えば、ユーザコントロールのタップ、またはタッチ画面上の基準発話の視覚化部のタップに応答して）音響再生を提供することができる。基準発話の音響表現を、オーディオファイルとしてメモリ12に事前に保存しておくこともできる。音響表現（例えば、再生）は、音響出力部15からユーザに提供可能であり、この音響出力部15は、コンピューティング装置の内部スピーカーまたは外部スピーカー（例えば、コンピューティング装置に有線接続または無線接続されるヘッドセット）に結合可能である。いくつかの実施形態では、基準発話の再生は、簡略化された視覚化部の表示に後続または先行する所定の期間の後に、または場合によって簡略化された視覚化部と同時になど、自動的に実施可能である。いくつかの実施形態では、基準発話の初回再生を、自動的に実施することができる。いくつかの実施形態では、ユーザが音響再生を命令することを可能にするユーザコントロールは、基準発話の視覚化部1004aであってもよいし、または基準発話を再生するように構成された別個のユーザコントロールを設けてもよい。次のステップに移行する前に、ユーザ1001（例えば、言語学習者）が所望する回数だけ基準発話の視覚化部をタップすることを可能にするように、アプリを構成することができ、装置は、例えばユーザによって命令された回数だけ基準発話を再生することができる。いくつかの実施形態では、基準発話のテキスト文字列1006を表示することもできる。説明したように、テキスト文字列1006は、発声された発話に関するいかなる韻律情報も有さないかもしれないが、その一方で、視覚化部1004aは、言語学習経験においてユーザを補助するための韻律情報を伝達することができる。いくつかの実施形態では、視覚化部1004aを表示することは、視覚化部のオブジェクトのアニメーションを表示することを含むことができ、このアニメーションは、発話の再生（学習者によって発声された発話および/または基準発話の再生）にリアルタイムで付随することができる。例えば、発話入力それぞれの分節（例えば、音節）が再生される際に、視覚化部の対応するオブジェクトを、その再生されている分節と実質的に同期させてアニメーション化することができる（例えば、新たに出現させる、強調表示する、既に表示されている場合には振動させる、点滅させる、サイズを変更する、軌道に沿って移動させるなどによって移動させることが可能であるか、またはその他のアニメーション化も可能である）。1つの具体的であるが非限定的な例として、アニメーションは、先行する分節と比較してより強さが大きくなっている分節（例えば、音節）（例えば、ストレスのかかった音節）に対応するオブジェクトを拡大すること、明るくすること、または強調表示することを含むことができる。別の具体的であるが非限定的な例では、視覚化部において、アクセントに起因するか、またはフレーズの語尾（例えば、発声された疑問文の語尾）にあるような、関連する分節のピッチパラメータの降下または上昇に対応する軌道に沿って、オブジェクトを移動させることができる。発話における韻律をリアルタイムでより忠実に表現していると思えることができるより豊かな視覚化部を提供するために、本明細書における任意のアニメーション例を組み合わせ使用することができる。本明細書で説明されているようなリアルタイムでの韻律表出のアニメーションは、学習者が新しい言語（または所与の言語の特定の方言）で発話するために発声およびリスニングの

10

20

30

40

50

練習をする際におけるユーザ体験を向上させる改善された学習ツールを提供することができる。

【 0 0 3 9 】

装置はさらに、ユーザ（例えば、言語学習者）が装置上で自身の発話を録音することを可能にするように構成されたユーザコントロール（例えば、録音アイコン 1 0 0 8）を表示することができる。図 1 0 B の画面ショット 1 0 0 2 - 2 に示されているように、ユーザ（例えば、言語学習者）は、このユーザコントロールを選択することができ（例えば、タッチ画面上のアイコンをタップする）、これに応答して、装置が録音モードに突入し、（例えば、装置に埋め込まれているか、または装置に通信可能に結合されている）マイクロフォンを使用してユーザの発話を録音するための、装置の録音機能が作動させられる。例えば、装置 1 0 において、プロセッサ 1 1 は、マイクロフォン入力部 1 4 を作動させることができ、これにより、マイクロフォン入力部 1 4 に結合された内部マイクロフォンまたは外部マイクロフォンが、言語学習者によって発話として生成された声の音圧を検出し、これにより、発話の録音が実施される。発話の録音は、一時的に（例えば、言語訓練セッションまたはその一部の持続時間の間）、または永続的に（例えば、ユーザによって明示的に削除されるまで）、図 1 のメモリ 1 2 のような装置のメモリに保存可能である。1 つの実施形態では、装置は、その後、ユーザ 1 0 0 1（例えば、言語学習者）の録音された発話を処理するために図 2 A の分節化プロセスを実行することができる。別の実施形態では、装置は、ユーザの録音された発話をリモートサーバに送信することができる。リモートサーバは、言語学習者の録音された発話に対して図 2 A の分節化プロセスを実行し、その録音された発話の分節化結果を装置に返送することができる。ユーザの録音された発話が分節化された後、装置は、ユーザの録音された発話の分節を表現するオブジェクト 1 0 0 3 - 1 , 1 0 0 3 - 2 , ~ 1 0 0 3 - n を含んでいる図像表現を作成するなどによって、録音された発話の視覚的表出 1 0 0 4 b を生成するためのプロセス（例えば、図 2 B のプロセス）を実行することができる。図 1 0 C の画面ショット 1 0 0 2 - 3 から見て取れるように、両方とも同じ視覚化プロセスを使用して生成される、基準発話の視覚化部 1 0 0 4 a と、ユーザの録音された発話の視覚化部 1 0 0 4 b とには違いが見られるが、この違いは、主として、発声された発話の内容（例えば、テキスト文字列）ではなく、発話のそれぞれ異なる 2 つの発声（1 つは基準、もう 1 つはユーザ）の韻律情報の違いに起因する可能性がある。このようにして、簡略化された視覚化部は、ユーザがネイティブの発話とユーザ（例えば、学習者）自身の発話との間の違いを容易に知覚することを可能にし、ユーザの発話学習プロセスを補助することができる。図 1 0 C にさらに示されているように、装置は、この例の視覚化部 1 0 0 4 b（例えば、オブジェクト 1 0 0 3 - 1 , 1 0 0 3 - 2 等の視覚化部）、または図 1 0 D にさらに示されているようなユーザの任意の後続する発声を、ユーザが保存することを可能にするように構成された追加的なユーザコントロール（例えば、録音アイコン）を、タッチ画面上のオブジェクトの図像表現と一緒に表示することができる。ユーザ命令（例えば、録音アイコン 1 0 1 0 のタップ）に応答して、またはいくつかの実施形態では視覚化部 1 0 0 4 b が生成されると自動的に、装置は、視覚化部 1 0 0 4 b をメモリ 1 2 に永続的に（例えば、ユーザによって明示的に削除されるまで）保存することができる。保存されたユーザの発声の視覚化部の分類、検索、レポート生成、および他の後続処理を行うことを可能にするために、ユーザの発声の視覚化部 1 0 0 4 b にタイムスタンプおよび/またはタグを付けることができる。これらの視覚化部をタグ付けして保存することにより、経時的に取得される保存された視覚化部と一緒に表示するなどによって、言語学習者の進捗を観察することが可能となる。ここでは、例えば、非ネイティブスピーカーが外国語を学習したい場合の言語学習の文脈で説明されているが、図 1 0 A ~ 1 0 D を参照しながら説明した実施形態のような本発明の実施形態は、他の目的で、例えば、演技のためのナレーションの練習のため、同じ言語の異なるアクセントまたは方言の学習のため、または任意の他の種類の発話練習または発話訓練のために使用可能である。本明細書で説明されている発話視覚化ツールの他の用途は、フレーズの発声を通じた自己啓発の練習であってよい。例えば、本明細書の視覚化技術を利用して

10

20

30

40

50

、習慣形成の練習またはツールを構築することができ、そこで、習慣形成プロセスの一部としてワードフレージングを使用することができる。

【 0 0 4 0 】

言語学習アプリに加えて、本明細書で説明されている視覚化技術のための他のユースケースも考えられる。例えば、本明細書で説明されている視覚化技術を中心にして、コミュニケーションアプリを構築することができる。いくつかの実施形態では、本明細書で説明されているプロセスによって生成される視覚化部は、他者と共有することができるユーザ生成コンテンツであってよい。1つのそのような例では、スマートフォンのテキストまたはビデオのメッセージングアプリのようなメッセージングアプリケーションを、視覚化機能と統合することができ、ここでは、メッセージングアプリを介して共有される任意の他の（例えば、テキスト、画像、ビデオ）メッセージの代わりに、またはそれらと組み合わせ、本明細書の任意の例に従って生成された発話視覚化部が提供される。これにより、特にテキストメッセージングの場合には、テキストだけでは伝達することができない情報（例えば、韻律情報）、例えば、発話されるメッセージの感情的なニュアンス、詳細等を伝達することが可能となる。

10

【 0 0 4 1 】

さらに、テキストのみによるコミュニケーション（例えば、テキストメッセージング）は、時として直接的すぎる、事実にすぎる、またはストレートすぎる可能性があり、効果的なコミュニケーションを促進しない可能性がある。そのような直接的かつ事実にコミュニケーションに感情的なニュアンスを吹き込むために、本明細書で説明されている視覚化技術を使用することができ、これにより、より効果的なコミュニケーションを提供することができる。このことを、テキストメッセージングだけでなく、（リモートでの）教育、コーチング、メンタリング、カウンセリング、セラピー、およびケアリングの分野にも適用することができる。教育という文脈では、本明細書の視覚化技術は、話者の発話能力に関する測定可能なデータを伝達し、この発話能力を経時的に追跡することができ、本明細書の技術に従って作成された視覚化部に関連するデータを使用して、練習および進捗を追跡することもできる。さらに、測定可能なデータを経時的に収集することができ、収集されたデータを種々異なる目的のために使用することができる。特に、学習者の練習データは、学習者自身、学習者を支援する教育者またはスタッフ、もしくは学習者に関連する他のユーザにとって有用であり得る。例えば、システムは、学習者の発話から作成された視覚化部におけるオブジェクトの個数をカウントすることによって発話の品質を定量的に分析することができる。分節（例えば、音節、音素等）を表現するそれぞれのオブジェクトは、物理的な筋肉練習の単位として見なすことができる。言語（例えば、英語）学習の練習では、学習者は、視覚化部におけるオブジェクトによって表現されるような、特定の個数（例えば、100万個）の分節の生成を達成することができ、その個数がカウントされる。本明細書で説明されている視覚化技術を実装する言語（例えば、英語）学習課程の1つの具体的であるが非限定的な例では、ユーザ（例えば、言語学習者）は、例えば毎日（または異なる頻度で、例えば週3回、週5回等）リスニングおよび発声の練習をする可能性がある。所与のそのような（例えば、毎日の）練習セッションは、特定の期間（例えば、ユーザに応じて15～30分）かかる可能性があり、したがって、ユーザは、1日当たり約15～30分（または異なる持続時間）を言語の練習に費やす可能性がある。練習セッションの間、ユーザは、それぞれが特定の個数の音節、例えば1フレーズ当たり8～9個の音節を有している特定の個数のフレーズ、例えば20～25個のフレーズを練習するよう求められる可能性がある。この具体例をさらに続けると、所与の練習セッションにおいてユーザがこの個数のフレーズを特定の回数、例えば14回繰り返した場合には、ユーザは、3000個を超える分節の発声を生成したこととなり、例えば、ユーザが毎日練習した場合には、100万単位を超える発声に相当することとなり、このような100万単位を超える発声は、マクロスケールでは（例えば、1年では）とても達成できそうにないかなりのチャレンジのように思えるが、1練習セッション当たりまたは1発声単位当たり分解すれば、言語を学習し始めたユーザにとってより身近に感じられ、したがって、

20

30

40

50

言語練習においてユーザを動機付けるために役立つ可能性がある。また、マクロレベルで（例えば、１年にわたって）生成された分節の総数をユーザに伝達することも、毎日の一歩一歩の練習がどのようにして経時的に蓄積され、発声／筋肉練習の大きな成果を達成することができるのかがユーザに示されることとなるので、ユーザの動機付けになる可能性がある。したがって、視覚化部におけるオブジェクトのカウントのように、視覚化部から取得することができる測定可能なデータであって、かつ視覚的なフィードバックを何ら得ることなくユーザが単に発声／練習しているだけでは利用できなかったであろう測定可能なデータは、発声練習の定性的な側面および定量的な側面の両方を分析するために有用であり得る。さらに、視覚化部におけるオブジェクトは、ユーザの行動分析のような他の目的にも有用であり得る。データサイエンス技術の分野における種々異なる技術を、（例えば、種々異なるように繰り返される発声および関連する視覚化部における）経時的に収集されたデータに対して個々におよび集合的に適用して、追加的な定性的情報および／または定量的情報を抽出することができる。

10

【 0 0 4 2 】

種々異なる他のアプリケーションでは、人物の発話を、現在の視覚化方法を介して包含および伝達される韻律情報に基づいてさらに特徴付けることができ、この情報を、例えば、ユーザのアバターまたは他の代理を作成するため、またはＡＩスピーカー（Googleホーム、アレクサ、Siri装置等）によって使用するために、他の装置、システム、またはプロセスによって使用することができ、これらは、ユーザのコミュニケーションを模倣するまたはより良好に理解するために、所与のユーザの韻律情報を利用することができる。また、視覚化技術は、本明細書では、図像オブジェクト（例えば、楕円形、長方形、または別の異なる形状のオブジェクト）をディスプレイ上に生成および表示することとして説明されているが、その一方で、他の例では、離散的な図像オブジェクトを含んでいる視覚化部の代わりに、適切な電子機器の離散的な発光素子（または発光素子の離散的なグループ）を順次に照明することもできる。いくつかの例では、米国特許第 9 2 1 8 0 5 5 号明細書（坂口ら）、米国特許第 9 9 4 6 3 5 1 号明細書（坂口ら）、および米国特許第 1 0 2 2 2 8 7 5 号明細書（坂口ら）に記載されているようなエンパセティックコンピューティング装置を使用して、本明細書で説明されている発話の視覚的表出を表出することができる。前述した特許は、如何なる目的であってもその全てが参照により本明細書に援用されるものとする。

20

30

【 0 0 4 3 】

図 1 1 A ～ 図 1 1 E は、本開示のさらなる実施形態による発話視覚化部を、発話のテキスト表現と組み合わせて提供する装置の画面キャプチャである。いくつかの実施形態では、図 1 1 A ～ 図 1 1 E の画面キャプチャに示されているようなユーザインターフェースを、携帯型コンピューティング装置（例えば、スマートフォン）のディスプレイによって生成して、このディスプレイ上に提供することができる。したがって、いくつかの例では、本開示による装置は、スマートフォンであってよく、このスマートフォンは、図 1 の装置 1 0 を実装しており、かつ図 1 の装置のディスプレイ画面 1 3 を実装するタッチ画面を有する。装置（例えば、スマートフォン）は、ユーザにテキストメッセージングサービスを提供するプログラム（例えば、テキストメッセージングアプリ）を実行するように構成可能である。テキストメッセージングアプリを、本開示による発話視覚化部によって拡張することができる。いくつかの実施形態では、（例えば、ユーザがテキストメッセージングアプリを使用しているときに）リアルタイムで録音された発話に対して視覚化を実施することができ、この録音された発話を、テキストに変換して視覚化部 1 1 0 4 と一緒にテキストメッセージングアプリを介して送信することができるか、または視覚化部 1 1 0 4 を、テキスト表現（例えば、ユーザのテキストメッセージ）の代わりに送信することができる。他の実施形態では、装置は、ユーザの発話の発声をモデル化するモデルを使用することができ、これにより、装置上で打ち込まれたテキストメッセージを視覚化して、この視覚化部を、ユーザ生成コンテンツとして他者と共有することができる。クラウドから取得可能であり、かつ／またはオプションとして装置 1 0 のメモリ 1 2 に保存可能である S V

40

50

E（またはそのコンポーネント）が搭載されたアプリケーション（「アプリ」）によって、拡張されたテキストメッセージングアプリケーションを実装することができる。

【0044】

図11Aでは、装置（例えば、スマートフォン）は、例えば拡張されたテキストメッセージングアプリがこの装置上で実行されているときに、メッセージインターフェース画面1102を表示するように構成されている。メッセージインターフェース画面1102は、ユーザがテキストメッセージを作成することを可能にする1つまたは複数のソフトコントロール1103（例えば、キーボードであるか、または音声メッセージを録音するための録音ボタンであり、この音声メッセージは、その後、装置によってテキストに変換される）のような、標準的なグラフィカルユーザインターフェース（GUI）コントロール要素（ソフトコントロールとも称される）を含むことができる。メッセージインターフェース画面1102は、メッセージが受信者に送信される前のメッセージのドラフトを表示するメッセージウィンドウ1105を表示することができる。メッセージインターフェース画面1102は、メッセージを打ち込むためのキーを含んでいるキーボードを表現するソフトコントロール1103を含むことができ、追加的にオプションとして、他のアプリケーション（アプリ）またはそれに関連するデータにアクセスするための1つまたは複数のソフトコントロール（例えば、アイコン1107）を含むことができる。いくつかの例では、メッセージインターフェース画面は、ユーザが画像、ビデオ、音楽、個人生体データ等のような種々異なるユーザ生成コンテンツを添付すること、および/または特定のアイコンに関連するアプリまたはその機能を作動させることを可能にするように構成された1つまたは複数のアイコン1107を表示することができる。拡張されたテキストメッセージングアプリにおいて、メッセージインターフェース画面1102は、本明細書の例に従って発話の視覚的表現を解析および生成することができる発話視覚化アプリ（SVA）のアイコン1107-1を追加的に含むことができる。図11Aに示されているような発話視覚化アプリアイコン1107-1が選択（例えば、タップ）されると、発話視覚化アプリが作動させられ、この発話視覚化アプリは、ユーザが本例に従って発話視覚化部1104を生成することを可能にするために、図11Bに示されているような、この発話視覚化アプリの独自のSVAインターフェースウィンドウ1109をテキストメッセージングアプリの内部に提供する。SVAインターフェースウィンドウ1109の一部として、装置（例えば、スマートフォン）は、ユーザが装置（例えば、スマートフォン）上で自身の発話を録音すること、および/または以前に録音された発話の視覚化部、またはユーザによって生成または受信されたテキストメッセージの視覚化部を生成することを可能にするアイコン1109-1を表示することができる。本例では、図11Cに示されているように、ユーザがタッチ画面上のアイコン1109-1をタップすると、装置は、録音モードに突入し、装置のマイクロフォンを使用して録音機能を作動させ、ユーザの発話を録音することができる。例えば装置10を参照すると、プロセッサ11は、音響入力部14を作動させることができ、これにより、音響入力部14に結合された内部マイクロフォンまたは外部マイクロフォンが、ユーザの発話によって生成された音波を検出し、検出された音波を発話入力（すなわち、発話波形または発話信号）として録音して、プロセッサ11に提供する。検出された発話の録音は、図1の装置のローカルメモリ12のような、装置10に通信可能に結合されたメモリに一時的または永続的に保存可能である。いくつかの実施形態では、ユーザは、発話を録音および変換するためのテキストメッセージングアプリの機能などを介して、発話視覚化アプリケーション（SVA）の外部で自身の発話を録音することができる。そのような場合には、SVAが作動させられると、ユーザは、別のアイコンをタップして、以前に録音された発話を検索し、以前に録音された発話の視覚化部1104を、SVAを介して生成することができる。装置（例えば、スマートフォン）は、本明細書の任意の例に従って前述したように発話の分節のためのオブジェクトを作成するなどにより、発話の視覚的表出を生成するための1つまたは複数のプロセスを実行することができる。図11Cに示されているように、装置は、オブジェクトの図像表現を、メッセージドラフトとしてこのオブジェクトの図像表現を確認することをユーザに促すメッセ

10

20

30

40

50

ージ確認アイコンと一緒にタッチ画面上に表示する。したがって、ユーザが、SVAインターフェースウィンドウ1109に表示された視覚化部に満足した場合には、ユーザは、アイコン（例えば、アイコン1109-2）をタップして、ユーザ生成コンテンツ（例えば、視覚化部1104）をテキストメッセージングアプリ（例えば、図11Dに示されているようなメッセージウィンドウ1105）に転送し、それにより、ここでは視覚化部1104の形態であるメッセージを、テキストメッセージングアプリのソフトコントロール（例えば、送信アイコン1103-s）を介して、図11Eのインターフェース画面1102eに示されているように受信者に送信することができる。受信者への送信は、無線伝送ネットワークを介して、例えば図1の装置10の無線送信器/受信機17を介して実施可能である。図11Eのインターフェース画面1102eにさらに示されているように、受信者は、テキストの形態で受信される従来のテキストメッセージの場合と同様に、ここでは視覚化部1104の形態で受信したメッセージ1111と相互作用（例えば、いいね、返信等）することができる。

10

【0045】

図12A~図12Dは、本開示の実施形態による、発話の生成された視覚的表現をそのタッチ画面上に含んでいるコミュニケーションシステムを提供する装置1200の画面キャプチャである。いくつかの実施形態では、図12A~図12Dの画面キャプチャに示されているようなユーザインターフェースを、携帯型コンピューティング装置（例えば、スマートフォン）のディスプレイによって生成して、このディスプレイ上に提供することができる。したがって、いくつかの例では、本開示による装置1200は、スマートフォンであってよく、このスマートフォンは、図1の装置10を実装しており、かつ図1の装置のディスプレイ画面13を実装するタッチ画面を有する。装置1200（例えば、スマートフォン）は、ユーザに視覚的メッセージングサービスおよび/またはテキストメッセージングサービスを提供するプログラム（例えば、メッセージングアプリ）を実行するように構成可能である。本開示によれば、メッセージングアプリは、本明細書の任意の例に従って（例えば、SVEによって）生成された発話視覚化部、および/または発話視覚化部を組み込んでいるコンテンツ、または少なくとも部分的に発話視覚化部に基づいているコンテンツを、ユーザが共有（例えば、送信および受信）することを可能にするように構成可能である。いくつかの実施形態では、メッセージングアプリは、クラウドに常駐するか、またはローカルに（例えば、装置10のメモリ12に）保存されているSVE（またはそのコンポーネント）と相互作用して、発話視覚化部を取得し、発話視覚化部を組み込んでいる関連するコンテンツ、または部分的に発話視覚化部に基づいている関連するコンテンツを生成する。いくつかの実施形態では、（例えば、ユーザがメッセージングアプリを使用しているときに）リアルタイムで録音された発話に対して視覚化を実施することができ、この録音された発話を、オプションとして、その関連するコンテンツ（例えば、アイコン1207A, 1208B, または1208D）と一緒にユーザに表示することができ、かつ/または受信側のユーザに送信することができる。

20

30

【0046】

図12A~12Dでは、装置（例えば、スマートフォン）1200は、例えばメッセージングアプリがこの装置1200上で実行されているときに、メッセージインターフェース画面1202を表示するように構成されている。図12A~図12Dは、ユーザがメッセージングアプリと相互作用してコンテンツを送受信しているときの、メッセージングインターフェース画面1202の種々異なるグラフィカルユーザインターフェース要素の例を示す。図12Aでは、メッセージングインターフェース画面1202は、送信者から受信したアイコン1207Aを含むメッセージウィンドウ1206Aを表示する。アイコン1207Aは、テキスト要素、図像要素、発話視覚化部要素、またはそれらの任意の組み合わせのような1つまたは複数の異なる種類のコンテンツ要素を含むことができる。図12Aのアイコン1207Aは、テキストメッセージ1208A、この例ではテキスト文字列“Sorry”と、送信者の発話に対応する発話視覚化部1209Aとを含み、例えば、この送信者の発話は、例えば自身の装置に“Sorry”という言葉を発声する送信者によって録音

40

50

可能であり、その後、コンテンツ（アイコン 1 2 0 7 A）が生成されて、受信側のユーザに送信される。この例のアイコン 1 2 0 7 A は、録音および視覚化された発話されたメッセージに基づいて送信者の装置によって選択された図像 1 2 1 0 A をさらに含む。メッセージングアプリは、多数の図像を保存するメモリ（例えば、ローカルメモリ 1 2 またはクラウド上のメモリ装置）と通信することができ、これらの図像の各々は、それぞれ異なるメッセージ、例えば“Sorry（ごめんね）”、“No problem（大丈夫です）”、“No worries（気にしないで）”、“Got it（了解）”、“Thanks（ありがとう）”、“Talk soon（またね）”のような一般的なメッセージに（例えば、ルックアップテーブルを介して）関連付けられている。いくつかの例では、同じまたは類似のアイコン（例えば、親指を立てることを含む図像）を、複数の異なるテキスト文字列（例えば、“Got it”または“No problem”）に 10 関連付けることができ、したがって、そのアイコンを選択して、それらの複数の異なるテキストメッセージのいずれかに関連するコンテンツに組み込むことができる。コンテンツ（例えば、アイコン 1 2 0 7 A）の図像（例えば、1 2 0 8 A）は、特定のテキストメッセージ（例えば、1 2 0 8 A）に典型的に関連する情報（例えば、感情）を視覚的に伝達することができ、したがって、メッセージングアプリによってテキストメッセージのみによってではなくコンテンツを介して伝えることは、ユーザ体験を豊かにすることができる。いくつかの例では、アイコン 1 2 0 7 A は、テキストメッセージ 1 2 0 8 A のユーザの発音に関する情報（例えば、メッセージが発話されたピッチ、速度等）を追加的に伝達 20 することができ、このことにより、コンテンツ作成者の心理状態（例えば、コンテンツ作成者の感情）に関する追加的な情報を送信者に伝達することができる。このようにして、例えば、従来のメッセージングアプリであれば取り込むことができなかった、または利用することができなかったユーザの発話に関する情報を伝えることによって、メッセージングサービスを向上させることができる。

【0 0 4 7】

ユーザがメッセージングアプリと相互作用すると、メッセージインターフェース画面 1 2 0 2 は、アプリとの相互作用を通じて作成された追加的な GUI 要素を表示するように更新される。例えば、図 1 2 B に示されているように、アイコン 1 2 0 7 B を含む第 2 のメッセージウィンドウ 1 2 0 6 B がメッセージインターフェース画面 1 2 0 2 に表示される。この例におけるアイコン 1 2 0 7 B は、装置 1 2 0 0 のユーザによって生成されたコンテンツを表現する。いくつかの例では、メッセージインターフェース画面 1 2 0 2 は、 30 ユーザが種々異なる他のユーザ生成コンテンツを添付すること、および/またはユーザの装置 1 2 0 0 に常駐するか、またはユーザの装置 1 2 0 0 に通信可能に結合された他のアプリまたはその機能を作動させることなど、他のアプリケーションと相互作用することを可能にするための種々異なるユーザコントロール（例えば、図 1 1 A のアイコン 1 1 0 7 のうちのいずれか 1 つまたは複数）を含むことができる。例えば、ここで図 1 2 C も参照すると、アイコン 1 1 0 7 のうちの 1 つは、ユーザが装置 1 2 0 0 の音声録音機能を作動させることを可能にすることができる。

【0 0 4 8】

図 1 2 C にさらに示されているように、メッセージインターフェース画面 1 2 0 2 は、例えば音声録音機能の作動に応答して、自身が録音した発話をユーザが視覚化することを 40 可能にするアイコンを表示することができ、オプションとしてこのアイコンを、別のメッセージウィンドウ 1 2 0 6 C（例えば、音声録音機能が作動させられると自動的に作成される）に表示することができる。メッセージングアプリでは、メッセージウィンドウ 1 2 0 6 C の内部に、またはメッセージインターフェース画面 1 2 0 2 の別の適切な場所に表示することができるアイコン 1 2 1 1 を表示することができ、このアイコン 1 2 1 1 は、ユーザによって選択されると、本明細書の例に従って（例えば、発話視覚化エンジン（SVE）を使用して）録音された発話の視覚的表現 1 2 0 9 C を生成するように構成されている。いくつかの実施形態では、例えばアイコン 1 2 1 1 のような単一のアイコンの選択に 50 応答して、メッセージングアプリの内部の録音機能を作動させることにより、発話視覚化機能も自動的に作動させることができる。さらに他の例では、メッセージングアプリの

内部の発話視覚化機能を作動させるために、アイコン（例えば、アイコン 1 1 0 7 - 1）を選択してもよく、この発話視覚化機能は、その後、ユーザが録音すること、および自身が録音した発話の視覚化部を生成することを可能にする。録音モードが作動させられるメカニズムにかかわらず、装置 1 2 0 0 は、（例えば、ユーザがアイコン 1 2 1 1 をタップすることに応答して）録音モードに突入し、これにより、装置 1 2 0 0 のマイクロフォン 1 2 0 1 を使用して録音機能を作動させ、ユーザの発話を録音することができる。例えば装置 1 0 を参照すると、プロセッサ 1 1 は、音響入力部 1 4 を作動させることができ、これにより、音響入力部 1 4 に結合された内部マイクロフォンまたは外部マイクロフォンが、ユーザの発話によって生成された音波を検出し、検出された音波を発話入力（すなわち、発話波形または発話信号）として録音して、プロセッサ 1 1 に提供する。検出された発話の録音は、図 1 の装置のローカルメモリ 1 2 のような、装置 1 0 に通信可能に結合されたメモリに一時的または永続的に保存可能である。いくつかの実施形態では、ユーザは、装置 1 2 0 0 の他の標準的な音声録音機能などを介して、メッセージングアプリの外部で自身の発話を録音することができる。そのような場合には、ユーザによってアイコン 1 2 1 1 が選択されると、メッセージングアプリは、以前に録音された発話を選択または検索するための GUI をユーザに提示することができ、その後、メッセージングアプリは、続いて、この以前に録音された発話の視覚化部 1 2 0 9 C を生成する。図 1 2 C の画面キャプチャでは、メッセージングアプリは、本明細書の例による視覚化部 1 2 0 9 C（例えば、振幅、ピッチ等のような発話の種々異なる特性を伝えるために色分けおよび配置することができる複数のオブジェクト 1 2 1 2 - 1 ~ 1 2 1 2 - 3）を生成している。いくつかの実施形態では、発話の視覚化部 1 2 0 9 C を、メッセージングインターフェース 1 2 0 2 の内部に（例えば、対応するアイコンが作成される前に一時的に）表示することができる。

10

20

【0049】

発話の視覚化に続いて、メッセージングアプリは、例えば図 1 2 D に示されているような、視覚化された発話に関連するコンテンツ（例えば、アイコン 1 2 0 7 D）を生成することができる。コンテンツ（例えば、アイコン 1 2 0 7 D）を、メッセージングインターフェース画面 1 2 0 2 に、例えばさらに別のメッセージウィンドウ 1 2 0 6 D に表示してもよいし、または視覚化された発話が表示されている場合には、この視覚化された発話と同じウィンドウ 1 2 0 6 C の内部に表示してもよい。いくつかの実施形態では、メッセージウィンドウ 1 2 0 6 D は、ユーザ生成コンテンツ（例えば、アイコン 1 2 0 7 D）を、このコンテンツが別のユーザに送信される前に表示するための確認ウィンドウであってよい。他のアイコン（例えば、アイコン 1 2 0 7 A および 1 2 0 7 B）と同様に、アイコン 1 2 0 7 D も、テキストメッセージ 1 2 0 8 D、図像 1 2 1 0 D、および/またはユーザの発話の視覚化部 1 2 0 9 D を含むことができる。この例では、アイコン 1 2 0 7 D は、共有されるべきユーザ生成コンテンツの内部に視覚化部 1 2 0 9 D を組み込んでいる（または含んでいる）。視覚化部 1 2 0 9 D を、人物のイラストであってよい図像 1 2 1 0 D に関連して、図像に描かれた人物の口に近接した位置などに適切に配置することができる。ユーザがユーザ生成コンテンツに満足すると、ユーザは、メッセージを別のユーザに送信するために構成されたアイコン 1 2 1 3 をタップすることができ、装置 1 2 0 0 は、これに応答して、意図された受信者にユーザ生成コンテンツ（例えば、アイコン 1 2 0 7 D）を送信することができ、その後、受信者に送信されたコンテンツ拡張メッセージのコピーをメッセージングアプリのメッセージウィンドウに表示することができる。図 1 2 D の例では、ユーザ生成コンテンツは、送信者からのメッセージに対する返信であってよく、したがって、ユーザ生成コンテンツを、メッセージ 1 2 0 7 A の送信者に提供することができる。ユーザがユーザ生成コンテンツ（例えば、アイコン 1 2 0 7 D）に満足していない場合には、ユーザは、発話を録音し直すことができ、このことにより、別の異なる視覚化部の列、ひいては別の異なる視覚化部 1 2 0 9 D を含むアイコン 1 2 0 7 D を作成することができる。

30

40

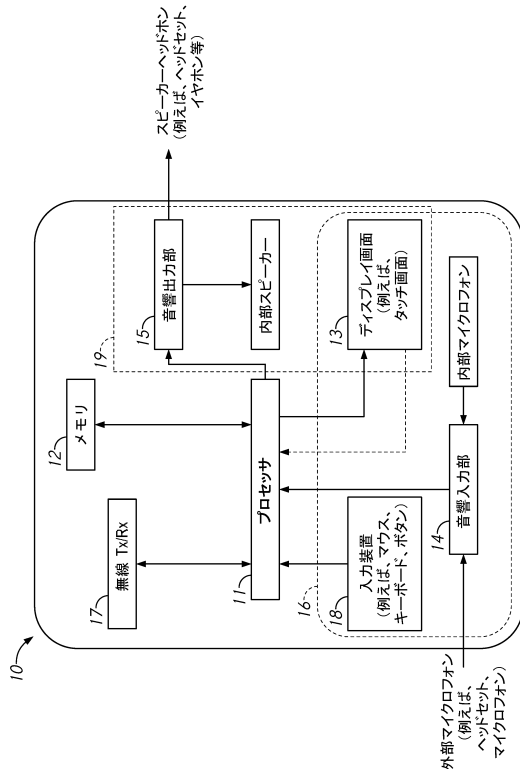
【0050】

50

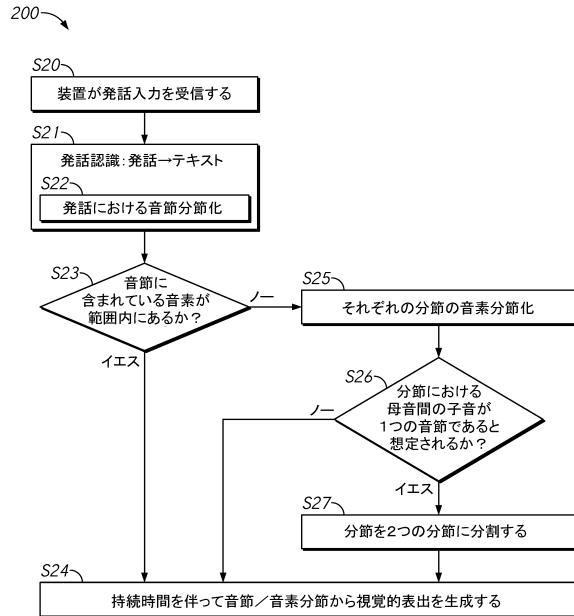
本発明は、上述した具体的な実施形態および例に限定されているわけではない。本発明を、説明された特定の組み合わせ以外の異なる組み合わせで具現化してよいことが想定されている。また、実施形態の具体的な特徴および態様の種々異なる組み合わせまたは組み合わせの構成要素を作成してよく、これらもなお、本発明の範囲内に含めてよいことが想定されている。開示された本発明の多様な様式を形成するために、開示されている実施形態の種々異なる特徴および態様を互いに組み合わせること、または互いに置き換えることができることが理解されるべきである。したがって、本明細書に開示されている本発明の少なくとも一部の範囲は、上述した特定の開示されている実施形態によって限定されるべきではないことが意図されている。

【図面】

【図 1】



【図 2 A】



10

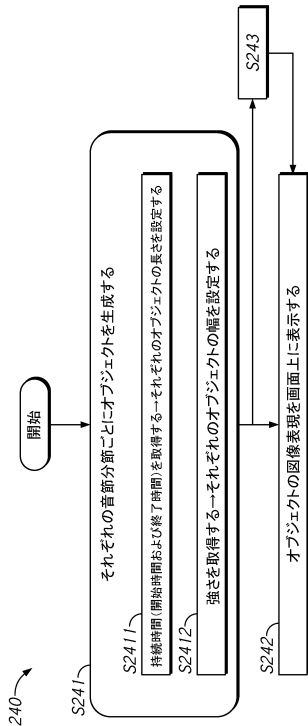
20

30

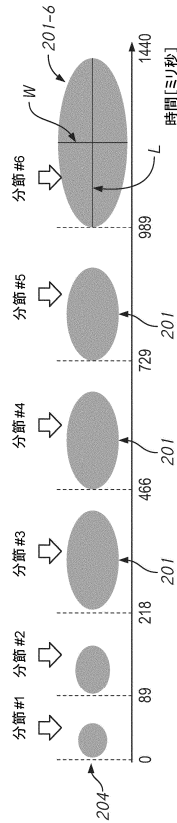
40

50

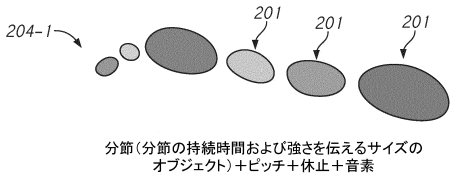
【図 2 B】



【図 2 C】

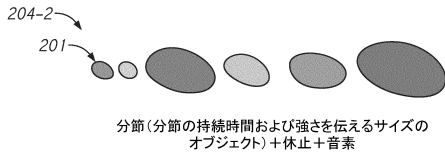


【図 2 D】



分節（分節の持続時間および強さを伝えるサイズのオブジェクト）+ピッチ+休止+音素

【図 2 E】



分節（分節の持続時間および強さを伝えるサイズのオブジェクト）+休止+音素

10

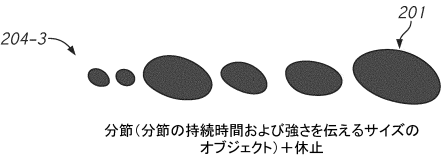
20

30

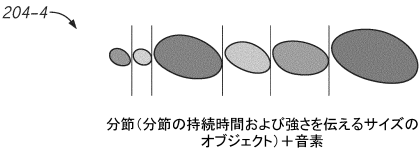
40

50

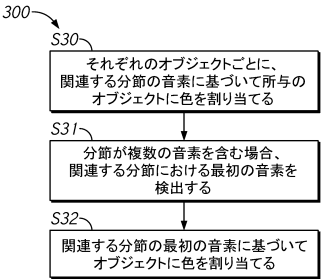
【図 2 F】



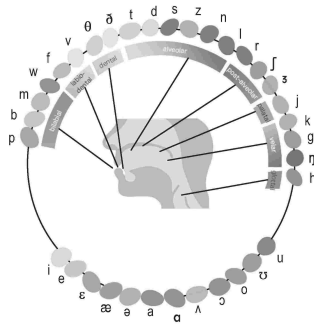
【図 2 G】



【図 3 A】



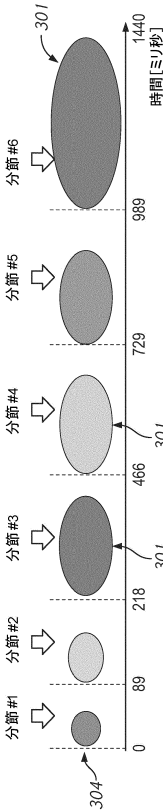
【図 3 B】



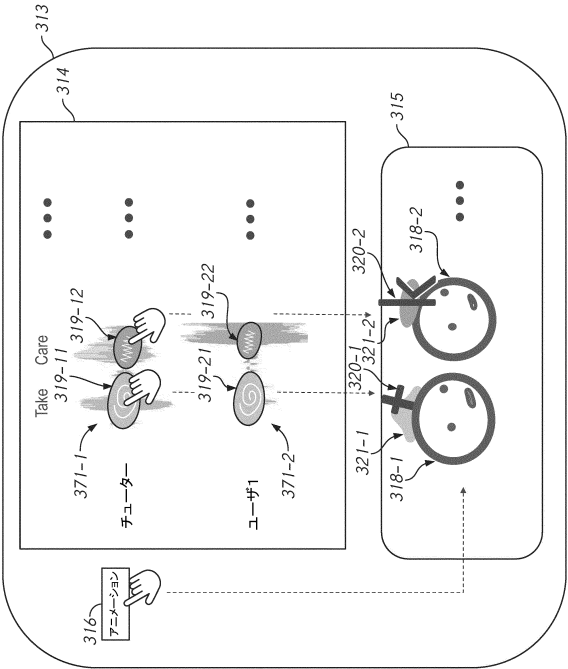
10

20

【図 3 C】



【図 3 D】

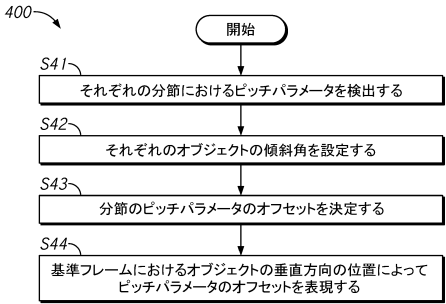


30

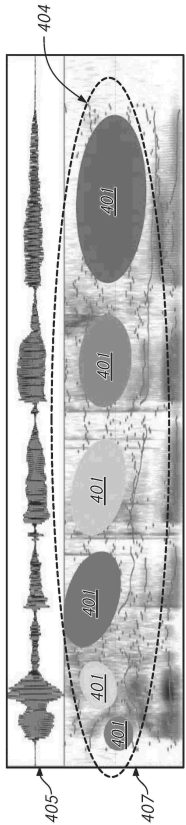
40

50

【図 4 A】



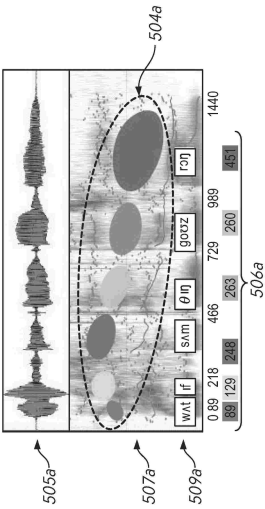
【図 4 B】



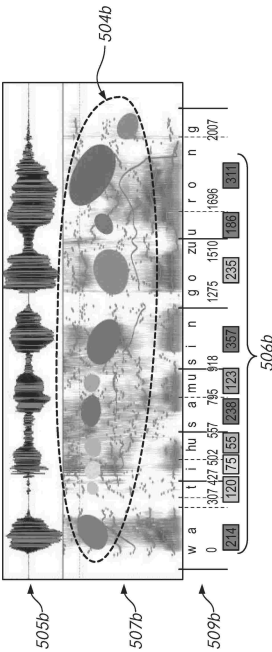
10

20

【図 5 A】



【図 5 B】

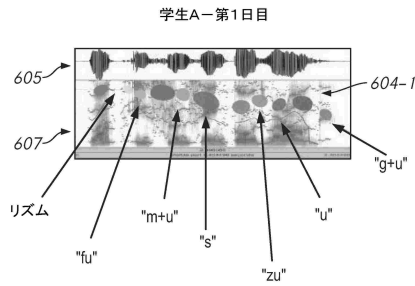


30

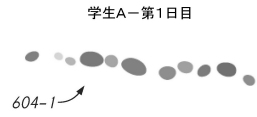
40

50

【図 6 A】

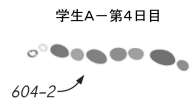


【図 6 B】

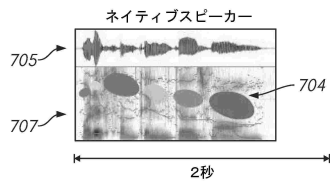


10

【図 6 C】

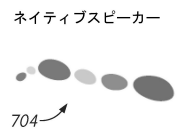


【図 7 A】

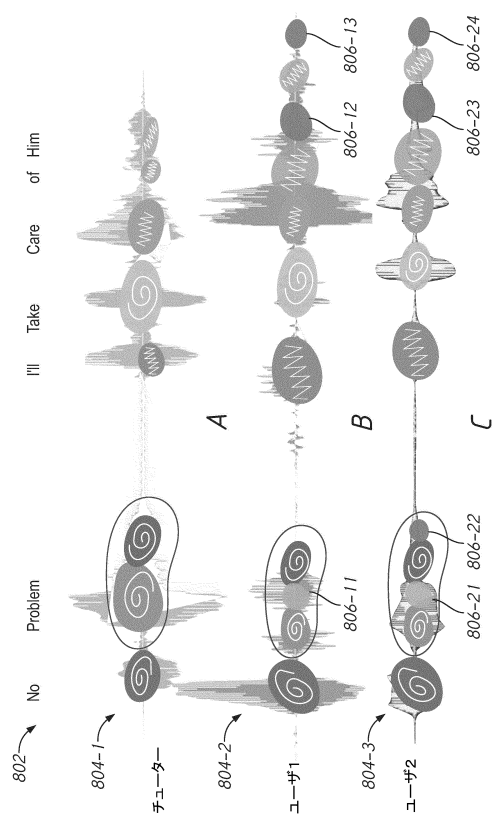


20

【図 7 B】



【図 8】

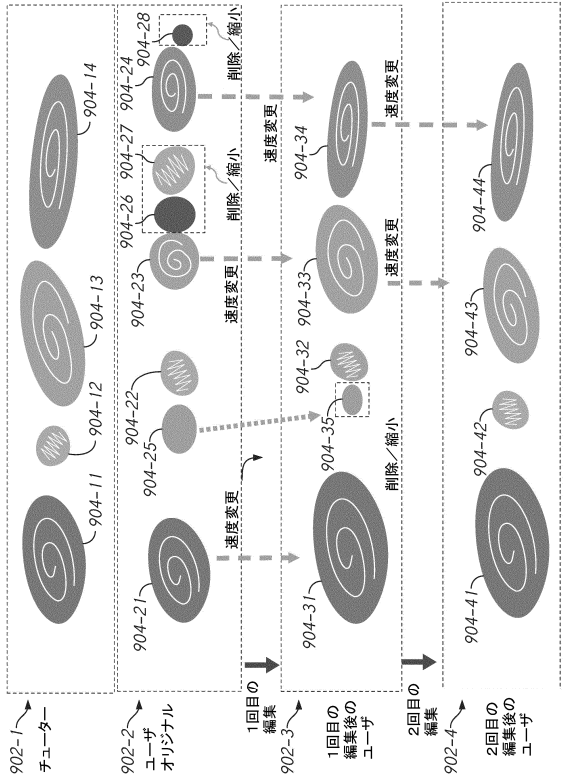


30

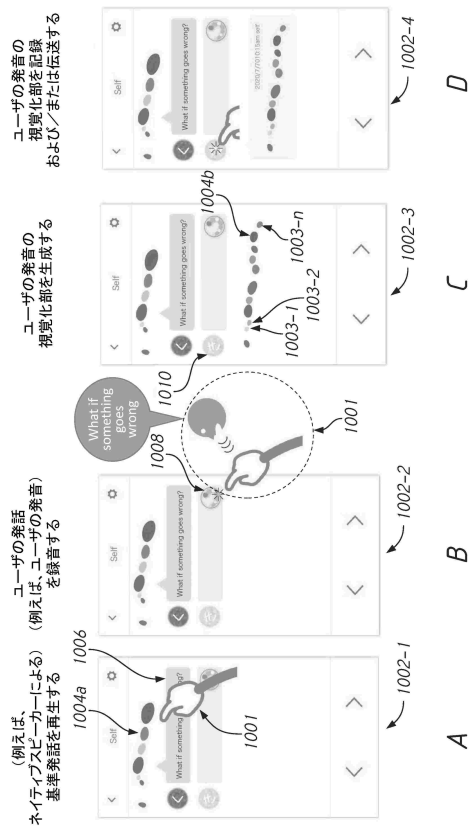
40

50

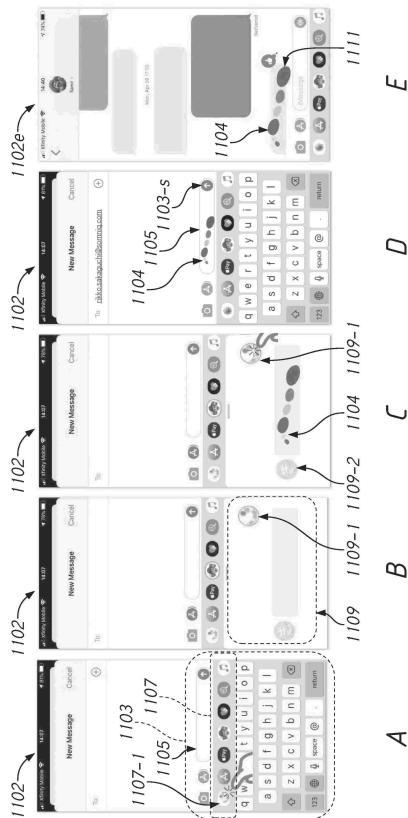
【図 9】



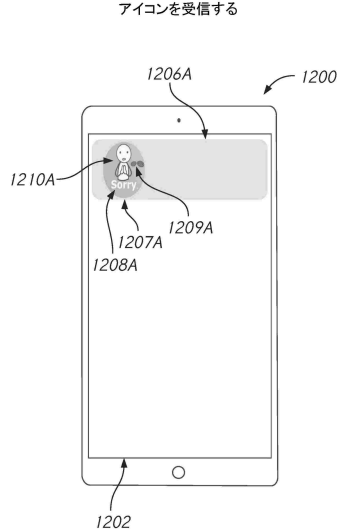
【図 10】



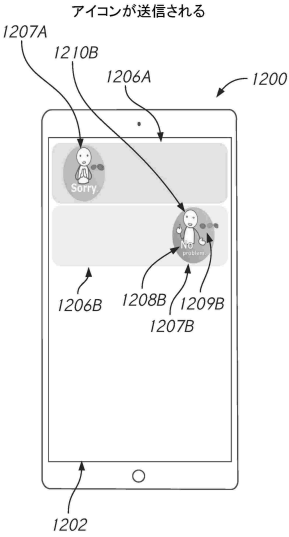
【図 11】



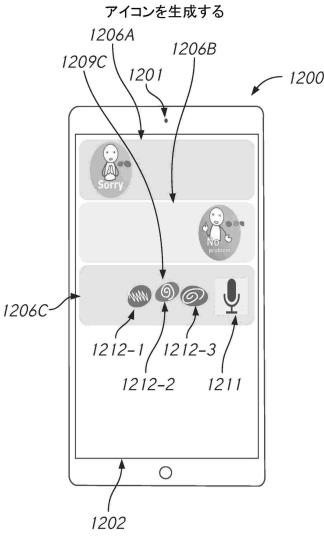
【図 12 A】



【図 1 2 B】

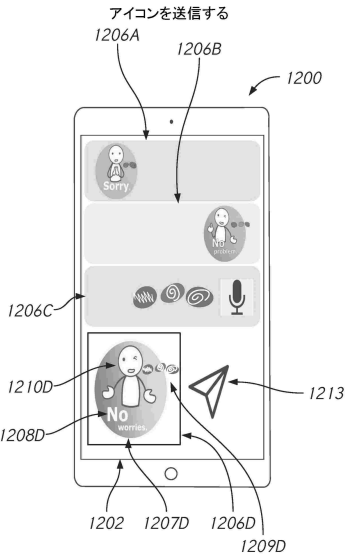


【図 1 2 C】



10

【図 1 2 D】



20

30

40

50

フロントページの続き

- (72)発明者 坂口 立考
アメリカ合衆国 カリフォルニア サニーヴェール サウス カシミア テラス 509
- (72)発明者 石川 英憲
アメリカ合衆国 カリフォルニア サニーヴェール サウス カシミア テラス 509
- 審査官 土井 悠生
- (56)参考文献 米国特許第06397185 (US, B1)
米国特許第06336089 (US, B1)
特開平10-133679 (JP, A)
特開2013-134475 (JP, A)
特開昭61-093484 (JP, A)
特開2003-186379 (JP, A)
- (58)調査した分野 (Int.Cl., DB名)
G10L 13/00 - 99/00
G09B 1/00 - 9/56
G09B 17/00 - 19/26