

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2013-161205

(P2013-161205A)

(43) 公開日 平成25年8月19日(2013.8.19)

(51) Int.Cl.

G06F 3/16 (2006.01)

F I

G06F 3/16 320D

テーマコード (参考)

審査請求 未請求 請求項の数 17 O L (全 19 頁)

(21) 出願番号 特願2012-21840 (P2012-21840)
 (22) 出願日 平成24年2月3日(2012.2.3)

(71) 出願人 000002185
 ソニー株式会社
 東京都港区港南1丁目7番1号
 (74) 代理人 100095957
 弁理士 亀谷 美明
 (74) 代理人 100096389
 弁理士 金本 哲男
 (74) 代理人 100101557
 弁理士 萩原 康司
 (74) 代理人 100128587
 弁理士 松本 一騎
 (72) 発明者 本間 文規
 東京都港区港南1丁目7番1号 ソニー株式会社内

最終頁に続く

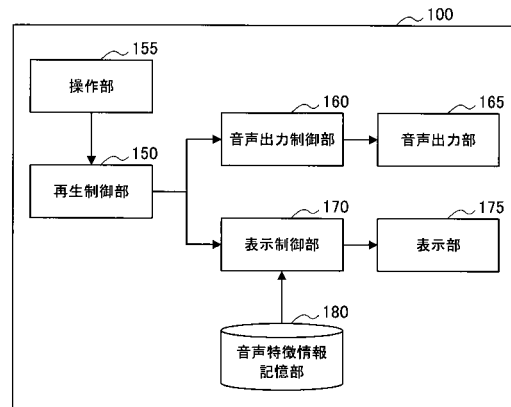
(54) 【発明の名称】 情報処理装置、情報処理方法、及びプログラム

(57) 【要約】

【課題】 音声の特徴を表示することができる情報処理装置、情報処理方法、及びプログラムを提供する。

【解決手段】 情報処理装置は、音声の出力を制御する音声出力制御部と、出力中の前記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御する表示制御部と、を有する。

【選択図】 図4



【特許請求の範囲】**【請求項 1】**

音声の出力を制御する音声出力制御部と、
出力中の前記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御する表示制御部と、
を備える、情報処理装置。

【請求項 2】

前記表示制御部は、出力中の前記音声の特徴が変化する点において、アニメーションオブジェクトのサイズ、形状、色、及び表示角度の少なくともいずれかを変化させる、
請求項 1 に記載の情報処理装置。

10

【請求項 3】

前記音声の前記特徴は、前記音声の音高、音長、及び音強の少なくともいずれかである、
請求項 1 に記載の情報処理装置。

【請求項 4】

前記アニメーションオブジェクトは、テキスト上に重畳して表示される、
請求項 1 に記載の情報処理装置。

【請求項 5】

前記テキスト上に重畳して表示される前記アニメーションオブジェクトは、出力中の音声位置に同期してサイズが変化するカーソルであり、前記カーソルの先端サイズは、前記音声の特徴が変化する点において前記形態が変化する、
請求項 4 に記載の情報処理装置。

20

【請求項 6】

前記アニメーションオブジェクトは、前記音声に対応するテキストである、
請求項 1 に記載の情報処理装置。

【請求項 7】

前記表示制御部は、出力中の前記音声の特徴に基づいて前記テキストの文字間隔を変化させる、
請求項 6 に記載の情報処理装置。

【請求項 8】

前記音声に対応するテキストは、動画コンテンツの字幕テキストである、
請求項 1 に記載の情報処理装置。

30

【請求項 9】

音声の出力を制御することと、
出力中の前記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御することと、
を含む、情報処理方法。

【請求項 10】

コンピュータを、
音声の出力を制御する音声出力制御部と、
出力中の前記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御する表示制御部と、
を備える、情報処理装置として機能させるためのプログラム。

40

【請求項 11】

前記表示制御部は、出力中の前記音声の特徴が変化する点において、アニメーションオブジェクトのサイズ、形状、色、及び表示角度の少なくともいずれかを変化させる、
請求項 10 に記載のプログラム。

【請求項 12】

前記音声の前記特徴は、前記音声の音高、音長、及び音強の少なくともいずれかである、
請求項 10 に記載のプログラム。

50

請求項 10 に記載のプログラム。

【請求項 13】

前記アニメーションオブジェクトは、テキスト上に重畳して表示される、
請求項 10 に記載のプログラム。

【請求項 14】

前記テキスト上に重畳して表示される前記アニメーションオブジェクトは、出力中の音声位置に同期してサイズが変化するカーソルであり、前記カーソルの先端サイズは、前記音声の特徴が変化する点において前記形態が変化する、
請求項 13 に記載のプログラム。

【請求項 15】

前記アニメーションオブジェクトは、前記音声に対応するテキストである、
請求項 10 に記載のプログラム。

【請求項 16】

前記表示制御部は、出力中の前記音声の特徴に基づいて前記テキストの文字間隔を変化させる、
請求項 15 に記載のプログラム。

【請求項 17】

前記音声に対応するテキストは、動画コンテンツの字幕テキストである、
請求項 10 に記載のプログラム。

10

20

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、情報処理装置、情報処理方法、及びプログラムに関する。

【背景技術】

【0002】

母国語以外の言語を学習しようとするとき、音声を聴くことにより発音を学習することがある。このとき音声に対応するテキストが表示されれば、ユーザは音声の内容を把握することが容易となる。例えば特許文献 1 には、映像に付加された字幕を中心に映像中の再生位置をサーチすることのできる再生装置が開示されている。この再生装置は、字幕に基づいて繰り返し再生を行うことができる。このため、聞き取り難い箇所を繰り返し再生することができ、学習効果が高められる。

30

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】特開平 09 - 115224 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかし、表示されたテキストを見ることによって、ユーザは音声の内容を把握することはできても、音声の特徴を把握することはできない。そこで、本開示では、音声の特徴を表示することの可能な、新規かつ改良された情報処理装置、情報処理方法、及びプログラムを提案する。

40

【課題を解決するための手段】

【0005】

本開示によれば、音声の出力を制御する音声出力制御部と、出力中の上記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御する表示制御部と、を有する情報処理装置が提供される。

【0006】

50

また、本開示によれば、音声の出力を制御することと、出力中の上記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御することと、を含む情報処理方法が提供される。

【0007】

また、本開示によれば、コンピュータを、音声の出力を制御する音声出力制御部と、出力中の上記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御する表示制御部と、を有する情報処理装置として機能させるためのプログラムが提供される。

【発明の効果】

【0008】

以上説明したように本開示によれば、音声の特徴を表示することができる。

【図面の簡単な説明】

【0009】

【図1】本開示の一実施形態に係る語学学習装置が表示する音声の特徴を示すアニメーションの概要を示す説明図である。

【図2】同実施形態に係る語学学習装置が表示する音声の特徴を示すアニメーションの概要を示す説明図である。

【図3】同実施形態に係る語学学習装置のハードウェア構成の一例を示すブロック図である。

【図4】同実施形態に係る語学学習装置の機能構成の一例を示すブロック図である。

【図5】同実施形態に係る語学学習装置の表示する表示画面の第1の例である。

【図6】同実施形態に係る語学学習装置の表示する表示画面の第2の例である。

【図7】同実施形態に係る語学学習装置の用いる音声特徴情報について説明するための説明図である。

【図8】同実施形態に係る語学学習装置の用いる音声特徴情報を含む字幕情報の一例を示す説明図である。

【図9】同実施形態に係る語学学習装置が音声特徴情報を用いて行う再生制御の一例を説明するための説明図である。

【図10】同実施形態に係る語学学習装置が音声の特徴に基づいて表示するアニメーションの一例を示す説明図である。

【図11】同実施形態に係る語学学習装置が音声の特徴に基づいて表示するアニメーションの他の一例を示す説明図である。

【図12】同実施形態に係る語学学習装置が音声の特徴に基づいて表示するアニメーションの他の一例を示す説明図である。

【図13】同実施形態に係る語学学習装置の第1の動作例を示すフローチャートである。

【図14】同実施形態に係る語学学習装置の第2の動作例を示すフローチャートである。

【図15】日本語と英語の音声の特徴を示す説明図である。

【図16】テキストに音声の波形情報を重畳表示した例を示す説明図である。

【図17】テキストを音声の長さにマッピングした表示例を示す説明図である。

【発明を実施するための形態】

【0010】

以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

【0011】

なお、説明は以下の順序で行うものとする。

1. 概要
2. ハードウェア構成例
3. 機能構成例
4. 表示画面例

10

20

30

40

50

- 5 . 音声特徴情報の例
- 6 . アニメーション例
- 7 . 動作例
- 8 . まとめ

【 0 0 1 2 】

< 1 . 概要 >

まず図 1 , 図 2 , 及び図 1 5 ~ 図 1 7 を参照しながら、本開示の一実施形態に係る語学学習サービスの概要について説明する。図 1 は、本開示の一実施形態に係る語学学習装置が表示する音声の特徴を示すアニメーションの概要を示す説明図である。図 2 は、同実施形態に係る語学学習装置が表示する音声の特徴を示すアニメーションの概要を示す説明図である。図 1 5 は、日本語と英語の音声の特徴を示す説明図である。図 1 6 は、テキストに音声の波形情報を重畳表示した例を示す説明図である。図 1 7 は、テキストを音声の長さにマッピングした表示例を示す説明図である。

10

【 0 0 1 3 】

本開示の一実施形態に係る語学学習サービスを提供する語学学習装置 1 0 0 は、音声の特徴をアニメーションによって示すことができる。なおここでは、日本人を対象として英語を学習する場合を想定して説明を行うが、本技術はかかる例に限定されず、母国語と異なる言語を学習するための学習サービスに適用することができる。

【 0 0 1 4 】

日本語と英語を比較すると、音声の音高（音の高低、以下ピッチという。）が異なるという特徴がある。日本語は比較的目的で見る文字と耳で聞こえる文字との一致度が高い言語であるのに対し、英語は目に見える文字と耳で聞こえる文字とが異なることが多い。例えば図 1 5 には、日本語の「緊急地震速報」の音声の特徴と、英語の「I'll inform you about our business hours.」の音声の特徴とが示される。日本語は、テキストの区切りがピッチの区切りと略一致し、ピッチが単語毎に小さな山を形成する。これに対し英語は、テキストの区切りがピッチの区切りとは一致せず、ピッチが文章全体で大きな山を形成する。このような違いがあるため、日本語を母国語とする人は、英語を聞き取ることが困難である。

20

【 0 0 1 5 】

そこで、英語のテキストと共に、テキストと対応する音声の特徴を表示することを提案する。ここで音声の特徴とは、例えば音声の音高（ピッチ）、音長（テンポ）、及び音強（インテンシティ）の少なくともいずれかを含んでよい。

30

【 0 0 1 6 】

音声の特徴を表示する方法としては、例えば図 1 6 に示されるように、音声の波形情報をテキストに重ねて表示することが考えられる。ところが、波形情報をテキストに重ねる場合には、波形を表示する分の高さを確保するために、テキスト間の行間が広めにとられる。このため、波形を表示しない場合の高さ h_1 と比較して、テキストに波形を重ねて表示した場合の高さ h_2 は大きくなってしまふ。

【 0 0 1 7 】

また英語のように音の長さや文字の長さが一致しない言語において、音の長さに合わせてテキストをマッピングすると、図 1 7 に示すようにテキストのレイアウトが崩れて読みづらくなってしまふ。また図示していないが、テキストの長さに合わせて音声を出力すると、当然ながら再生スピードが一定にならず聞き取りづらい。

40

【 0 0 1 8 】

このように、静的な表示によって音声の特徴を示そうとすると、様々な不都合が生じることがある。そこで本開示では、アニメーションを用いて音声の特徴を表示することを提案する。例えば図 1 及び図 2 に音声の特徴を示すアニメーションの一例が示される。

【 0 0 1 9 】

例えば「Then he'll ask, "Can you make more?"」という文章に対して付加されるアニメーションとし

50

ては、図1に示されるようにカーソル表示が挙げられる。このカーソルは、いわゆるカラオケにおいて歌詞のうち現在歌う部分と同期してカーソル位置が移動する場合と同様に、音声の出力位置に合わせて移動する。図1の上図には、「he」の部分の音声出力されているときのカーソルの状態が示される。また中図には、「ask」の部分の音声出力されているときのカーソルの状態が示される。また下図には、「you」の部分の音声出力されているときのカーソルの状態が示される。このように音声の出力位置とカーソル位置が同期される。これにより、音声のテンポがカーソル位置によって示される。

【0020】

またここで表示されるカーソルの形態は、図2に示されるように変化してもよい。例えば音声のピッチ又はインテンシティに応じてカーソル先端部のサイズが変化してもよい。図2のパターン1には、文章末で音声のピッチ又はインテンシティが下がる場合のアニメーション例が示される。音声のピッチ又はインテンシティが下がる時点において、カーソル先端部のサイズが縮小される。また図2のパターン2には、文章末で音声のピッチ又はインテンシティが上がる場合のアニメーション例が示される。音声のピッチ又はインテンシティが上がる時点において、カーソル先端部のサイズが拡大される。また図2のパターン3には、文章の途中で音声のピッチ又はインテンシティが上がった後元に戻る場合のアニメーション例が示される。音声のピッチ又はインテンシティが上がる時点において、カーソル先端部のサイズが拡大された後、元のカーソルサイズに戻される。

10

【0021】

上述の通り、出力される音声の特徴をテキスト全てについて静的に示す場合と比べて、“出力中の音声”の特徴をアニメーション(アニメーションオブジェクトの形態変化)で示す本技術によれば、限られたスペースで効果的に音声の特徴をユーザに示すことができる。このような語学学習サービスを実現するための語学学習装置100について、以下に説明する。

20

【0022】

<2. ハードウェア構成例>

まず図3を参照しながら、本開示の一実施形態に係る語学学習装置100のハードウェア構成の一例について説明する。図3は、同実施形態に係る語学学習装置のハードウェア構成の一例を示すブロック図である。

【0023】

語学学習装置100は、例えばアプリケーションを実行することによって本開示の一実施形態に係る語学学習サービスを提供することのできる情報処理装置の一例である。例えば語学学習装置100は、携帯電話、PC(Personal Computer)、映像処理装置、ゲーム機器、家電機器、音楽再生装置などの情報処理装置であってよい。

30

【0024】

(語学学習装置100)

語学学習装置100は、例えばCPU(Central Processing Unit)105と、RAM(Random Access Memory)110と、不揮発性メモリ115と、表示装置120と、タッチセンサ125と、音声出力装置130とを主に有する。

40

【0025】

CPU105は、演算処理装置および制御装置として機能する。CPU105は、各種プログラムに従って語学学習装置100内の動作全般を制御する。またCPU105は、マイクロプロセッサであってよい。RAM110は、CPU105が実行するプログラムや、プログラムの実行において適宜変化するパラメータ等を一時記憶する。不揮発性メモリ115は、CPU105が使用するプログラムや演算パラメータなどを記憶する。

【0026】

表示装置120は、表示部の一例であり、例えば液晶ディスプレイ(LCD: Liquid Crystal Display)装置、有機ELディスプレイ(OELD: Organic ElectroLuminescence Display)装置、ブラウ

50

ン管 (CRT: Cathode Ray Tube) ディスプレイ装置などであってよい。

【0027】

タッチセンサ125は、操作部の一例であり、表示装置120上に重畳して設けられる。タッチセンサ125は、位置入力装置であり、表示画面における位置情報に基づいて、操作を受付けることができる。

【0028】

音声出力装置130は、音声を出力するスピーカである。この音声出力装置130は、例えば再生されるコンテンツに含まれる音声を出力することができる。また音声出力装置130は、再生されるコンテンツに含まれるテキストデータから生成された合成音声を出力することもできる。

10

【0029】

以上、ここで示された語学学習装置100のハードウェア構成は一例であり、本技術はかかる例に限定されない。語学学習サービスを実現するためのハードウェア構成については、本実施形態を実施する時々の技術レベルなどに応じて、適宜、利用する構成を変更することができる。

【0030】

<3. 機能構成例>

次に、図4を参照しながら語学学習装置100の機能構成の一例について説明する。図4は、同実施形態に係る語学学習装置の機能構成の一例を示すブロック図である。

20

【0031】

語学学習装置100は、再生制御部150と、操作部155と、音声出力制御部160と、音声出力部165と、表示制御部170と、表示部175と、音声特徴情報記憶部180との機能を主に有する。

【0032】

再生制御部150は、コンテンツの再生を制御する機能を有する。再生制御部150は、操作部155から供給される操作情報に基づいて動作することができる。再生制御部150は、操作情報に基づいて、再生するコンテンツを選択することができる。また再生制御部150は、音声出力制御部160及び表示制御部170を制御することによってコンテンツの再生を制御することができる。

30

【0033】

操作部155は、ユーザが所望の操作を行うための入力装置であり、例えばタッチセンサ125であってよい。操作部155は、ユーザの操作に基づいた操作情報を生成して再生制御部150に共有することができる。操作部155は、例えばユーザが、再生するコンテンツを選択する操作、コンテンツ中の再生位置を選択する操作、及び再生に関する各種のパラメータ(例えば再生音量および再生速度など)を設定する操作を行うと、操作に応じた操作情報を生成することができる。

【0034】

音声出力制御部160は、再生制御部150の制御に従って音声出力部165が出力する音声を制御することができる。音声出力制御部160は、再生制御部150の指定するコンテンツ中の指定された再生箇所に対応する音声の出力を制御する。音声出力制御部160は、指定されたコンテンツ中に音声データが含まれる場合には、この音声データの出力を制御することができる。また音声出力制御部160は、指定されたコンテンツのテキストから合成音声を生成し、生成した合成音声の出力を制御することもできる。

40

【0035】

音声出力部165は、音声出力制御部160の制御に従って音声を出力する機能を有する。音声出力部165は、音声出力装置130であってよい。なおここでは語学学習装置100が音声出力装置130を有することとしたが、本技術はかかる例に限定されない。例えば音声出力部165は、外部の音声出力装置に音声を出力するインタフェースであってもよい。

50

【 0 0 3 6 】

表示制御部 170 は、再生制御部 150 の制御に従って表示部 175 の出力する表示画面の内容を制御することができる。表示制御部 170 は、再生制御部 150 の指定するコンテンツ中の指定された再生箇所を表示するように表示部 175 の表示画面の内容を制御することができる。表示制御部は、音声出力部 165 が出力中の音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御することができる。ここでアニメーションオブジェクトは、例えば音声の内容を示すテキスト上に重畳して表示されるオブジェクトであってよい。またアニメーションオブジェクトは、例えば音声の内容を示すテキストに含まれる文字であってよい。表示制御部 170 は、アニメーションオブジェクトの形態を変化させることにより、アニメーションで音声の特徴を示すことができる。ここでアニメーションオブジェクトの形態とは、例えば形状、サイズ、色、表示角度などであってよい。

10

【 0 0 3 7 】

表示部 175 は、表示制御部 170 の制御に従って表示画面をユーザに提供する機能を有する。表示部 175 は、表示装置 120 であってよい。ここでは語学学習装置 100 が表示装置 120 を有することとしたが、本技術はかかる例に限定されない。例えば表示部 175 は、外部の表示装置に表示画面を出力させるためのインターフェースであってよい。

【 0 0 3 8 】

音声特徴情報記憶部 180 は、音声特徴情報を記憶する機能を有する記憶部の一例である。音声特徴情報記憶部 180 は、例えば不揮発性メモリ 115 であってよい。音声特徴情報の詳細については後述されるが、音声の特徴を示す情報であり、例えば音声が所定の特徴を有する点をコンテンツ中において特定するための情報が含まれてよい。例えば、音声特徴点とは、音声のピッチ、テンポ、又はインテンシティが変化する点であってよい。

20

【 0 0 3 9 】

以上、本実施形態に係る語学学習装置 100 の機能の一例を示した。上記の各構成要素は、図 3 に示されたハードウェア構成により実現されてもよいし、汎用的な部材や回路が用いられてもよい。また上記の各構成要素は、各構成要素の機能に特化したハードウェアにより構成されていてもよい。また、各構成要素の機能を、CPU (Central Processing Unit) などの演算装置がこれらの機能を実現する処理手順を記述した制御プログラムを記憶した ROM (Read Only Memory) や RAM (Random Access Memory) などの記憶媒体から制御プログラムを読み出し、そのプログラムを解釈して実行することにより行ってもよい。従って、本実施形態を実施する時々の技術レベルに応じて、適宜、利用する構成を変更することが可能である。

30

【 0 0 4 0 】

なお、上述のような本実施形態に係る語学学習装置 100 の各機能を実現するためのコンピュータプログラムを作成し、パーソナルコンピュータ等を実装することが可能である。また、このようなコンピュータプログラムが格納された、コンピュータで読み取り可能な記録媒体も提供することができる。記録媒体は、例えば、磁気ディスク、光ディスク、光磁気ディスク、フラッシュメモリなどである。また、上記のコンピュータプログラムは、記録媒体を用いずに、例えばネットワークを介して配信してもよい。

40

【 0 0 4 1 】

< 4 . 表示画面例 >

次に、図 5 及び図 6 を参照しながら、本実施形態に係る語学学習装置 100 の表示する表示画面例について説明する。図 5 は、同実施形態に係る語学学習装置の表示する表示画面の第 1 の例である。図 6 は、同実施形態に係る語学学習装置の表示する表示画面の第 2 の例である。

【 0 0 4 2 】

50

ここでは、映画などの動画像コンテンツを用いた語学学習サービスを提供する第1の表示画面と、テキストコンテンツに対して選択した箇所の音声を出力する第2の表示画面とを例に挙げて説明をする。

【0043】

(第1の表示画面)

図5に示される第1の表示画面は、映画コンテンツを利用した語学学習サービスを提供する表示画面の一例である。図5に示される第1の表示画面は、コンテンツ表示領域11と、字幕表示領域12と、学習記録表示領域13と、コンテンツリスト表示領域14と、辞書表示領域15と、関連動画表示領域16とを含むことができる。字幕表示領域12は、アクティブ字幕表示領域を含むことができる。現在出力中の音声は、このアクティブ字幕表示領域に表示される。このアクティブ字幕表示領域には、後に詳述されるアニメーションオブジェクトが表示される。

10

【0044】

(第2の表示画面)

図6に示される第2の表示画面は、テキストデータのうち指定された箇所を音声出力する語学学習サービスを提供する表示画面の一例である。ここでは、インターネット上に公開されたHTMLコンテンツ中のテキストのうち指定された箇所の音声を出力することとしたが、本技術はかかる例に限定されない。例えばHTMLコンテンツに限らず、テキストデータを含むコンテンツ全般に対して本技術は適用されてよい。

20

【0045】

< 5 . 音声特徴情報の例 >

次に、図7～図9を参照しながら、本開示の一実施形態に係る語学学習装置100の用いる音声特徴情報の一例について説明する。図7は、同実施形態に係る語学学習装置の用いる音声特徴情報について説明するための説明図である。図8は、同実施形態に係る語学学習装置の用いる音声特徴情報を含む字幕情報の一例を示す説明図である。図9は、同実施形態に係る語学学習装置が音声特徴情報を用いて行う再生制御の一例を説明するための説明図である。

【0046】

映画のような動画像コンテンツの字幕及び音声データを再生する第1の表示画面の例を挙げると、音声特徴情報は、例えば図7に示されるフォーマットを有する。音声特徴情報は、例えばCommandIDと、音声の特徴点を特定するための時間と、それぞれの特徴点を区別するためのセパレータとを含んでよい。例えばCommandIDの一例としては、音高(ピッチ)が大きくなる点を示すPU、音高が小さくなる点を示すPD、テンポが変化する点を示すSPなどが挙げられる。記述例としては、PU24868; SP25058; 等が挙げられる。PU24868; は、24868ミリ秒の時点においてピッチが大きくなることを示す。またSP25058; は、25058ミリ秒の時点においてテンポが変わることを示す。

30

【0047】

このような音声特徴情報は、例えば図8に示されるように、字幕情報の一部として含まれていてもよい。例えば字幕情報は、字幕IDと、字幕表示時間と、音声特徴情報と、字幕テキストとを含むことができる。このような字幕情報を参照して、このコンテンツの再生を制御する場合の制御方法の一例が図9に示される。

40

【0048】

ここでは、字幕ID9の字幕再生時間が「00:00:34,448」から「00:00:36,575」であり、この字幕に対応する音声に対して、「SP34348; SP35568; SP35578; SP36188」の4つの音声特徴情報が付加されている場合を例に挙げて説明する。この字幕の字幕テキストは、「I wasn't talking about China anymore, I was talking about me.」である。このとき、語学学習装置100は、カーソルの表示をこの音声特徴情報に基づいて制御することができる。

【0049】

50

この音声特徴情報は、34348, 35568, 35578, 36188の4つの時点においてテンポが変化することを示す。従ってこの4つの時点は、字幕テキスト中の「I wasn't talking about China anymore,」の開始位置及び終了位置、及び「I was talking about me.」の開始位置及び終了位置に相当すると考えられる。そこで、語学学習装置100は、これらの情報を用いてカーソル速度を割り当てることができる。具体的にこの例について言えば、「I wasn't talking about China anymore,」の開始時間を34348ミリ秒、終了時間を35568ミリ秒、そして「I was talking about me.」の開始時間を35578ミリ秒、終了時間を36188ミリ秒とすることができる。

【0050】

<6. アニメーション例>

次に、図10～図12を参照しながら、本開示の一実施形態に係る語学学習装置100の表示するアニメーションオブジェクトの表示例について説明する。図10は、同実施形態に係る語学学習装置が音声の特徴に基づいて表示するアニメーションの一例を示す説明図である。図11は、同実施形態に係る語学学習装置が音声の特徴に基づいて表示するアニメーションの他の一例を示す説明図である。図12は、同実施形態に係る語学学習装置が音声の特徴に基づいて表示するアニメーションの他の一例を示す説明図である。

【0051】

表示制御部170は、例えば図1、図2、又は図10～図12に示されるアニメーションを用いることによって、音声の特徴を表示することができる。図10を参照すると、文字の形態を変化させることによって音声の特徴を示すアニメーションオブジェクトの一例が示される。ここでは、表示された文字を音声の特徴に基づいて変化させることによってアニメーション表現を実現することができる。例えば文字のサイズが拡大又は縮小されてよい。また文字の表示角度が回転されてもよい。また文字の一部（例えば頂部）が変形されてもよい。ここではアニメーションのパターンを表すグラフが複数示される。アニメーションオブジェクトは、このグラフに従って形態が変化されてよい。また図示されていないが、音声の特徴に応じて文字の色が変化されてもよい。表示制御部170は、音声特徴情報に示された特徴点においてアニメーションオブジェクトの形態を変化させることができる。

【0052】

また図11には、図2に示したカーソルの移動及び先端サイズ変化によるアニメーションオブジェクトに、文字サイズを変化させることによるアニメーションオブジェクトを組合せた一例が示される。例えばテキストデータ中に含まれる、音声の特徴点において、カーソルの先端部のサイズ及び文字のサイズが拡大又は縮小されてよい。

【0053】

また図12に示されるように、文字の間隔によって音声の特徴を示してもよい。テキスト全体を音声のテンポに合わせて表示すると、図17に示したように全体のレイアウトが崩れて読みづらくなり、また文字の間に音声のテンポに応じた空間が必要であるため、テキストを等間隔で詰めて表示する場合と比較して、表示領域が大きくなる。ところが、ここでは音声を出力している箇所のテキストを音声のテンポに合わせて表示し、出力している箇所以外のテキストについては詰めて表示される。またこのとき、出力中の音声に対応するテキストが拡大して表示領域が不足した分は、出力中の箇所以外のテキストが潰れて表示されてよい。これにより、テキストを等間隔で詰めて表示した場合と同じ大きさの表示領域で、出力中の箇所のテキストが音声のテンポに合わせて文字間隔が拡大して表示される。

【0054】

<7. 動作例>

次に図13及び図14を参照しながら、本開示の一実施形態に係る語学学習装置100の動作例について説明する。なお、ここでは図5に説明した映画などの動画像コンテンツを用いた語学学習サービスを提供する第1の表示画面を提供する場合の第1の動作例と、図6に説明したテキストコンテンツに対して選択した箇所の音声を出力する第2の表示画

10

20

30

40

50

面を提供する第2の動作例とについてそれぞれ説明する。図13は、同実施形態に係る語学学習装置の第1の動作例を示すフローチャートである。図14は、同実施形態に係る語学学習装置の第2の動作例を示すフローチャートである。

【0055】

(第1の動作例)

まず図13を参照しながら、第1の動作例について説明する。再生制御部150は、コンテンツ再生中であるか否かを判断する(S100)。すなわち、コンテンツの再生が停止されるまで、以下に説明する動作は繰り返される。

【0056】

そしてコンテンツ再生中であると判断されると、再生制御部150は、現在の再生時間を読み込む(S105)。再生制御部150は、音声出力制御部160及び表示制御部170に、再生時間に対応する音声又は表示画面を出力するように制御する。ここで表示制御部170は、現在の再生時間から次に表示する字幕と、字幕時間を読み込む(S110)。また表示制御部170は、音声特徴情報を読み込む(S115)。

10

【0057】

そして表示制御部170は、現在の再生時間に音声特徴点があるか否かを音声特徴情報に基づいて判断する(S120)。そして現在の再生時間に音声特徴点がある場合には、表示制御部170は、音声特徴情報に基づいたアニメーションを表示させる(S125)。

【0058】

かかる動作により、映画などの動画像コンテンツにおいて、音声に特徴がある箇所、例えばピッチ又はインテンシティが所定の閾値以上変化する箇所、或いはテンポが変化する箇所において、字幕上にアニメーションが表示される。ここで表示されるアニメーションは、アニメーションオブジェクトの形態の変化により実現される。アニメーションオブジェクトは、テキスト上に重畳されるカーソルであってもよい。或いはアニメーションオブジェクトは、テキスト自体であってもよい。また形態の変化とは、サイズ、形状、色、配置などの変化であってもよい。これにより、ユーザは、視覚的に音声の特徴を把握することができる。またアニメーションによって、「現在出力されている」音声の特徴だけを表示させることによって、表示領域を有効に利用することができる。

20

【0059】

(第2の動作例)

次に図14を参照しながら、第2の動作例について説明する。再生制御部150は、アプリケーション実行中であるか否かを判断する(S200)。すなわち、アプリケーションの実行が停止されるまで、以下に説明する動作は繰り返される。

30

【0060】

そしてアプリケーション実行中であると判断されると、再生制御部150は、操作部155がテキスト上のタッチ操作を検知したか否かを判断する(S205)。テキスト上でタッチ操作が検知されると、再生制御部150は、タッチ箇所のテキストに対応する音声データを特定する(S210)。再生制御部150は、特定した箇所の音声データを出力するように音声出力制御部160を制御し、これに応じた表示画面を表示するよう表示制御部170を制御する。ここで音声出力制御部160は、特定した音声データの出力を開始する(S215)。なおここでは予め音声データが存在するものとしたが、本技術はかかる例に限定されない。例えば上述の通り、タッチ箇所のテキストが特定されると、音声出力制御部160は、特定されたテキストから合成音声を生成してもよい。

40

【0061】

そして表示制御部170は、音声特徴情報を読み込む(S220)。また表示制御部170は、現在の音声出力時間を読み込む(S225)。そして表示制御部170は、現在の音声出力時間に音声特徴点があるか否かを判断する(S230)。そして現在の音声出力時間に音声特徴点があるときには、表示制御部170は、音声特徴情報に基づいたアニメーションを表示させる(S235)。一方、現在の音声出力時間に音声特徴点がない場合に

50

は、このステップ S 2 3 5 の処理は省略される。次に再生制御部 1 5 0 は、特定された音声データを最後まで出力したか否かを判断する (S 2 4 0)。そしてまだ最後まで音声データが出力されていない場合には、ステップ S 2 2 5 に戻って処理が繰り返される。一方、音声データが最後まで出力された場合には、ステップ S 2 0 0 に戻って処理が繰り返される。

【 0 0 6 2 】

< 8 . まとめ >

以上説明したように、本開示の一実施形態に係る語学学習装置 1 0 0 によれば、音声の特徴を表示画面に効果的に示すことができる。アニメーションを用いることにより、コンテンツに含まれる音声全ての特徴を同時に示す場合と比較して、出力中の音声の特徴が効果的に示される。また出力中の音声の特徴のみを表示するため、限られた表示領域において、出力中の音声以外の箇所については通常が表示状態とすることができ、可読性も向上する。

10

【 0 0 6 3 】

また上記では添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、特許請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

20

【 0 0 6 4 】

例えば、本技術を適用する情報処理装置は、音声の特徴を表示する表示制御機能を有する装置全般に適用することができる。上記では、映画コンテンツの字幕上にカーソル表示する例と、HTMLコンテンツのテキスト上にアニメーション表示する例とを示したが、かかる例に限定されない。映画コンテンツの字幕上で音声の特徴をアニメーション表示する実施例によれば、映画コンテンツを語学学習の教材として利用することができる。またコンテンツのテキスト上に音声の特徴をアニメーション表示する実施例によれば、一般的なあらゆるテキストを語学学習の教材として利用することができる。なおハードウェアとしては、上記ではタッチセンサを有するタブレット端末又はスマートフォンを例に挙げたが、かかる例に限定されない。あらゆる情報処理装置に適用することができる。

30

【 0 0 6 5 】

また、本明細書において、フローチャートに記述されたステップは、記載された順序に沿って時系列的に行われる処理はもちろん、必ずしも時系列的に処理されなくとも、並列的に又は個別に実行される処理をも含む。また時系列的に処理されるステップでも、場合によっては適宜順序を変更することが可能であることは言うまでもない。

【 0 0 6 6 】

なお、以下のような構成も本開示の技術的範囲に属する。

(1)

音声の出力を制御する音声出力制御部と、

出力中の前記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御する表示制御部と、
を備える、情報処理装置。

40

(2)

前記表示制御部は、出力中の前記音声の特徴が変化する点において、アニメーションオブジェクトのサイズ、形状、色、及び表示角度の少なくともいずれかを変化させる、前記 (1) に記載の情報処理装置。

(3)

前記音声の前記特徴は、前記音声の音高、音長、及び音強の少なくともいずれかである、
前記 (1) または (2) のいずれかに記載の情報処理装置。

(4)

50

前記アニメーションオブジェクトは、テキスト上に重畳して表示される、
前記(1)～(3)のいずれかに記載の情報処理装置。

(5)

前記テキスト上に重畳して表示される前記アニメーションオブジェクトは、出力中の音声位置に同期してサイズが変化するカーソルであり、前記カーソルの先端サイズは、前記音声の特徴が変化する点において前記形態が変化する、
前記(4)に記載の情報処理装置。

(6)

前記アニメーションオブジェクトは、前記音声に対応するテキストである、
前記(1)～(5)のいずれかに記載の情報処理装置。

10

(7)

前記表示制御部は、出力中の前記音声の特徴に基づいて前記テキストの文字間隔を変化させる、
前記(6)に記載の情報処理装置。

(8)

前記音声に対応するテキストは、動画コンテンツの字幕テキストである、
前記(1)～(7)のいずれかに記載の情報処理装置。

(9)

音声の出力を制御することと、

出力中の前記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御することと、
を含む、情報処理方法。

20

(10)

コンピュータを、

音声の出力を制御する音声出力制御部と、

出力中の前記音声の特徴に基づいて形態が変化するアニメーションオブジェクトの表示を制御する表示制御部と、
を備える、情報処理装置として機能させるためのプログラム。

(11)

前記表示制御部は、出力中の前記音声の特徴が変化する点において、アニメーションオブジェクトのサイズ、形状、色、及び表示角度の少なくともいずれかを変化させる、
前記(10)に記載のプログラム。

30

(12)

前記音声の前記特徴は、前記音声の音高、音長、及び音強の少なくともいずれかである、
前記(10)または(11)のいずれかに記載のプログラム。

(13)

前記アニメーションオブジェクトは、テキスト上に重畳して表示される、
前記(10)～(12)のいずれかに記載のプログラム。

(14)

前記テキスト上に重畳して表示される前記アニメーションオブジェクトは、出力中の音声位置に同期してサイズが変化するカーソルであり、前記カーソルの先端サイズは、前記音声の特徴が変化する点において前記形態が変化する、
前記(13)に記載のプログラム。

40

(15)

前記アニメーションオブジェクトは、前記音声に対応するテキストである、
前記(10)～(14)のいずれかに記載のプログラム。

(16)

前記表示制御部は、出力中の前記音声の特徴に基づいて前記テキストの文字間隔を変化させる、

50

前記(15)に記載のプログラム。
(17)

前記音声に対応するテキストは、動画コンテンツの字幕テキストである、
前記(10)～(16)のいずれかに記載のプログラム。

【符号の説明】

【0067】

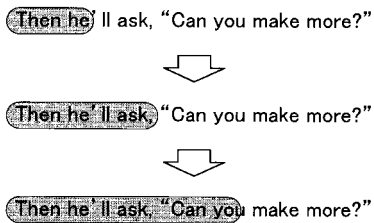
- 100 語学学習装置
- 105 CPU
- 110 RAM
- 115 不揮発性メモリ
- 120 表示装置
- 125 タッチセンサ
- 130 音声出力装置
- 150 再生制御部
- 155 操作部
- 160 音声出力制御部
- 165 音声出力部
- 170 表示制御部
- 175 表示部
- 180 音声特徴情報記憶部

10

20

【図1】

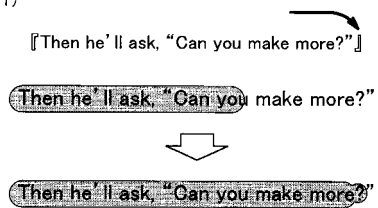
『Then he'll ask, "Can you make more?』



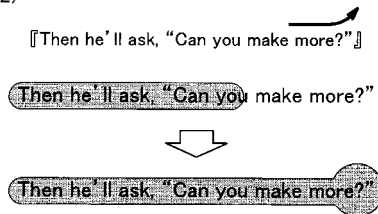
音声の出力位置とカーソル位置を同期

【図2】

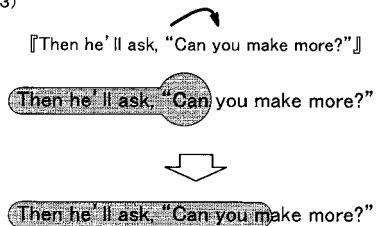
(パターン1)



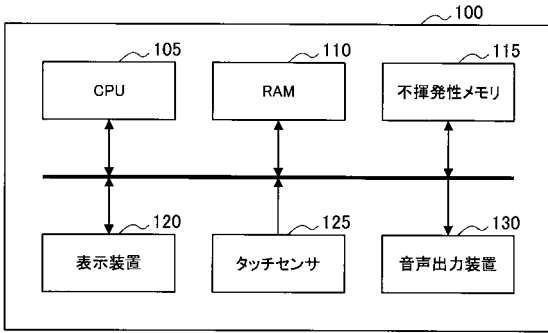
(パターン2)



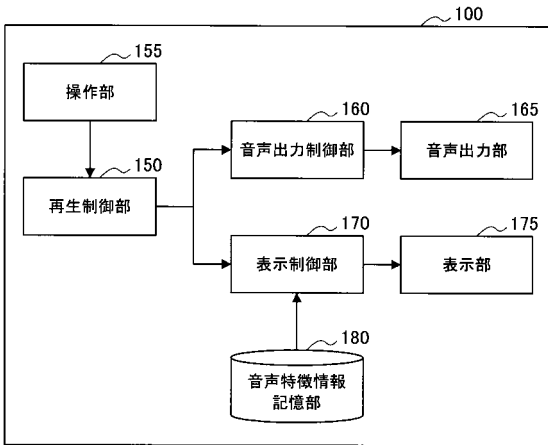
(パターン3)



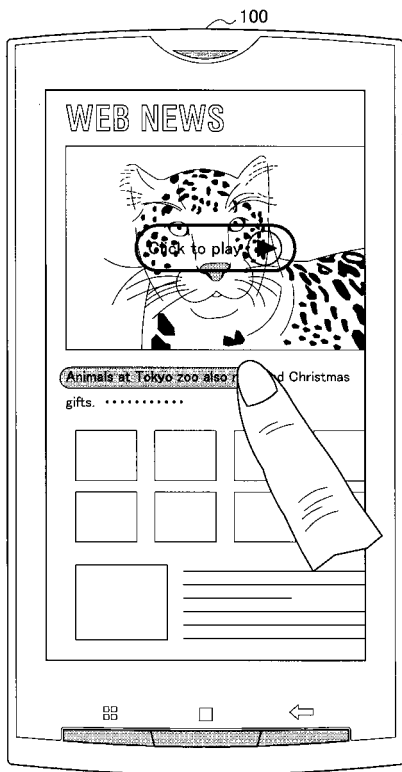
【 図 3 】



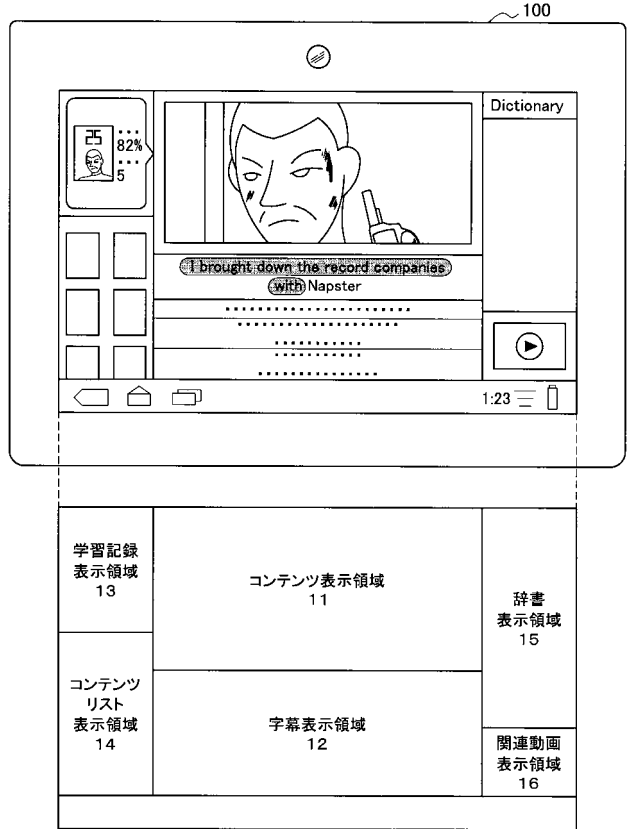
【 図 4 】



【 図 6 】



【 図 5 】



【 図 7 】

音声特徴情報のフォーマット例
CommandID+時間(ミリ秒)+セパレータ

CommandID(例)

- PU: 音高が大きくなる (pitch up)
- PD: 音高が小さくなる (pitch down)
- SP: テンポが変わる (カーソル速度の変化点)

記述例

- PU24868: 24868msecで音高が上がる
- SP25058: 25058msecでテンポが変わる

【 図 8 】

字幕例

1
00:00:17.798 _> 00:00:20.130
[SALA]PU17858;PD20138;
Did you know there are more people
with genius IQs living China
⇒字幕ID
⇒字幕表示時間
⇒音声特徴情報
⇒字幕テキスト

2
00:00:20.233 _> 00:00:22.258
[SALA]SP20178;PU22178
than there are people of any kind
living in the United States?

3
00:00:22.336 _> 00:00:24.133
[SALA]PU22228;PD23348;SP23358;SP23648
-That can't possibly be true.
-It is.

4
00:00:24.237 _> 00:00:25.261
[SALA]PU23728;PU24818;
What would account for that?

5
00:00:25.339 _> 00:00:28.331
[SALA]PU24868;PU25058;SP25068;SP25258;SP25268;PU26448;PD26458;SP27938;
Well, first, an awful lot of people
live in China, but here's my question.

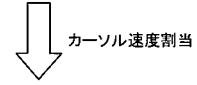
6
00:00:28.442 _> 00:00:30.569
[SALA]PU28388;SP30478;
How do you distinguish yourself
in a population of people

7
00:00:30.644 _> 00:00:32.475
[SALA]PU30538;SP31948;
who all got 1600 on their SATs?

8
00:00:32.579 _> 00:00:34.342
[SALA]PU32068;SP33848;SP33868;PD34278;
-I didn't know they take SATs in China.
-They don't.

【 図 9 】

9
00:00:34.448 _> 00:00:36.575
[SALA]SP34348;SP35568;SP35578;SP36188;
I wasn't talking about China anymore,
I was talking about me.



カーソル表示開始時間 (msec)	カーソル表示終了時間 (msec)
34348	35568
35578	36188

"I wasn't talking about China anymore,"
"I was talking about me."

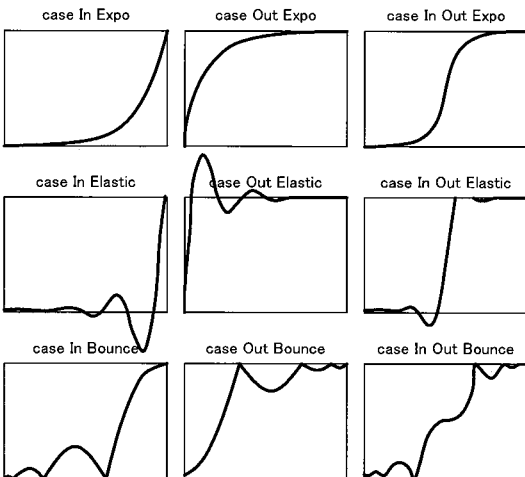
【 図 1 0 】



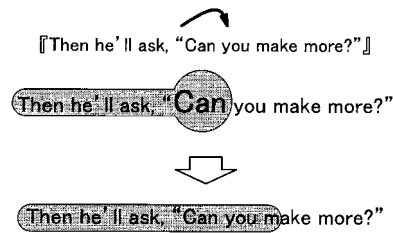
拡大縮小

回転

頂点変形

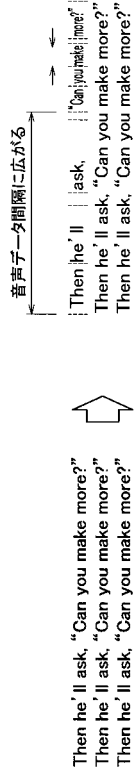


【 図 1 1 】

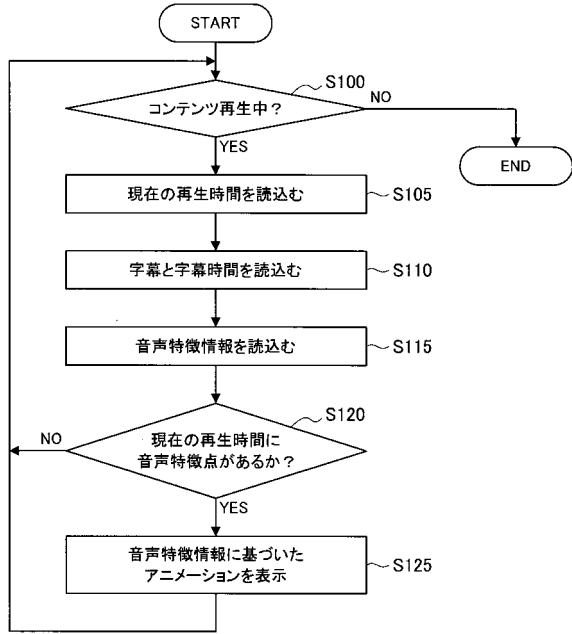


Then he'll ask, "Can you make more?"

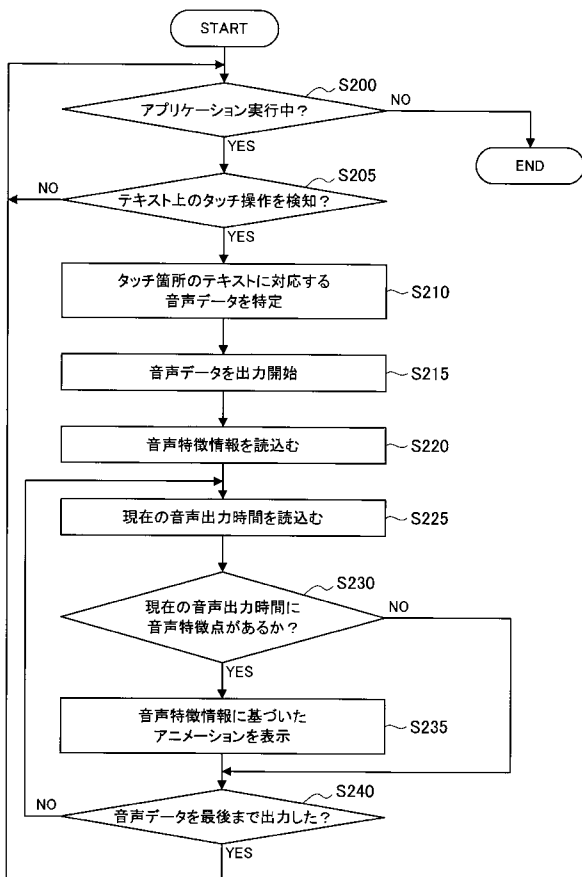
【 図 1 2 】



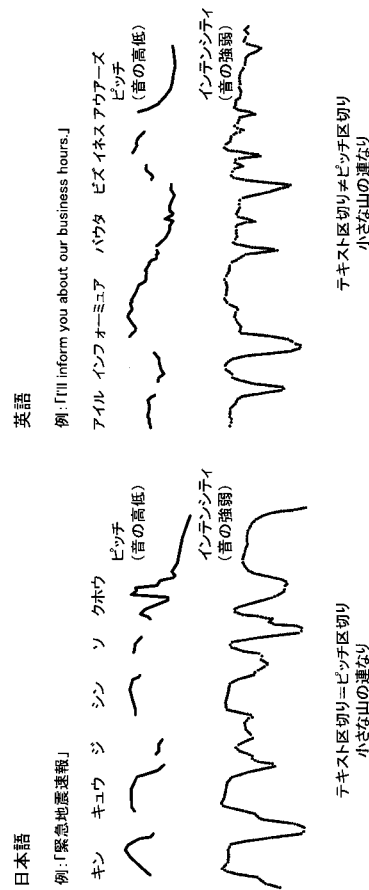
【 図 1 3 】



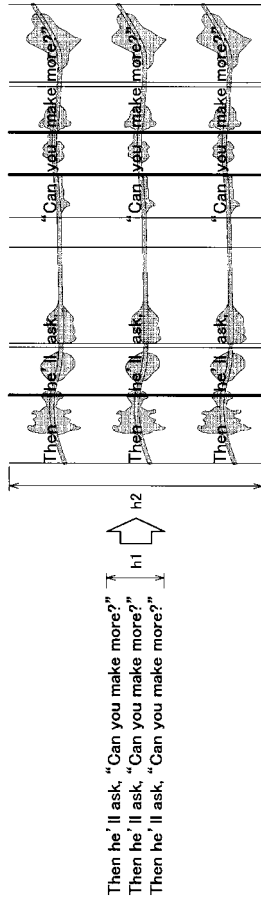
【 図 1 4 】



【 図 1 5 】



【 図 16 】



【 図 17 】



Then he'll ask, "Can you make more?"
 Then he'll ask, "Can you make more?"
 Then he'll ask, "Can you make more?"



Then he'll ask,
 Then he'll ask,
 Then he'll ask,

Can you make more?
 Can you make more?
 Can you make more?

フロントページの続き

- (72)発明者 奥村 泰史
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 宮下 健
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 梨子田 辰志
東京都港区港南1丁目7番1号 ソニー株式会社内