

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4627154号
(P4627154)

(45) 発行日 平成23年2月9日(2011.2.9)

(24) 登録日 平成22年11月19日(2010.11.19)

(51) Int.Cl.		F I			
G 1 0 L	15/10	(2006.01)	G 1 0 L	15/10	5 0 0 N
G 1 0 L	11/00	(2006.01)	G 1 0 L	11/00	4 0 2 H
G 1 0 L	19/00	(2006.01)	G 1 0 L	19/00	3 1 2 E
G 0 6 N	3/00	(2006.01)	G 0 6 N	3/00	5 5 0 E

請求項の数 8 (全 20 頁)

(21) 出願番号	特願2004-172615 (P2004-172615)	(73) 特許権者	301022471
(22) 出願日	平成16年6月10日(2004.6.10)		独立行政法人情報通信研究機構
(65) 公開番号	特開2005-352151 (P2005-352151A)		東京都小金井市貫井北町4-2-1
(43) 公開日	平成17年12月22日(2005.12.22)	(74) 代理人	100130111
審査請求日	平成19年6月7日(2007.6.7)		弁理士 新保 斉
		(72) 発明者	村田 真樹
			東京都小金井市貫井北町4-2-1 独立行政法人情報通信研究機構内
		(72) 発明者	一井 康二
			神奈川県横須賀市久里浜6-10-2-405
		(72) 発明者	岡井 隆弘
			東京都文京区音羽1-26-13-304
		審査官	毛利 太郎

最終頁に続く

(54) 【発明の名称】 人間の感情状態に応じた音楽出力装置及び音楽出力方法

(57) 【特許請求の範囲】

【請求項1】

人間の発する音声及び活動に伴って生じる活動音に対して、当該人の感情状態に適し、該音声及び該活動音の背景音楽を選択して出力可能な音楽出力装置であって、感情状態学習用テーブルと、
 該感情状態学習テーブルに基づいて第1の機械学習モデルにより音響特徴量と感情状態との関連を取得して状態判断テーブルに学習結果を保存する感情状態学習手段と、音楽選択学習用テーブルと、
 該音楽選択学習テーブルに基づいて第2の機械学習モデルにより選択する音楽と感情状態との関連を取得して音楽選択テーブルに学習結果を保存する音楽選択学習手段と、
 該音声及び該活動音を取得し、音響情報として入力する音響情報入力手段と、
 該音響情報から該音響情報の音響スペクトルに係る特徴量を抽出する音響特徴量抽出手段と、
 該音響特徴量を入力して、該状態判断テーブルに基づいて音響情報に係る人間の感情状態を該第1の機械学習モデルにより判断する状態判断手段と、
 判断された該感情状態を入力して、該音楽選択テーブルに基づいて該第2の機械学習モデルにより音楽を選択する音楽選択手段と、
 該音楽を背景音楽として出力する音楽出力手段と
 を備え、
 該音響スペクトルに係る特徴量として、強度、テンポ、及び周波数を全て用いる

10

20

ことを特徴とする人間の感情状態に応じた音楽出力装置。

【請求項 2】

前記感情状態学習用テーブルを、前記音声及び活動音を生じさせる人間の特徴別に用意して学習し、

前記状態判断手段において、判断時に選択された特徴別の学習結果を用いて音響情報に係る当該人間の感情状態を判断する

請求項 1 に記載の人間の感情状態に応じた音楽出力装置。

【請求項 3】

前記状態判断テーブルにおいて、

同一の音響特徴量に対して、複数の感情状態の確度が定義される

請求項 1 又は 2 に記載の人間の感情状態に応じた音楽出力装置。

10

【請求項 4】

前記機械学習モデルが、

最大エントロピーモデル、サポートベクトルマシンモデル

のいずれかである

請求項 1 ないし 3 のいずれかに記載の人間の感情状態に応じた音楽出力装置。

【請求項 5】

前記音楽出力装置が、電話回線との接続手段を備え、

前記音声情報入力手段が、電話回線における通話音声を取得して入力すると共に、

前記音楽出力手段が、電話回線における通話音声に音楽を挿入する

請求項 1 ないし 4 のいずれかに記載の人間の感情状態に応じた音楽出力装置。

20

【請求項 6】

人間の発する音声及び活動に伴って生じる活動音に対して、当該人の感情状態に適し、該音声及び該活動音の背景音楽を選択して出力可能な音楽出力方法であって、

感情状態学習手段により、予め備えた感情状態学習テーブルに基づいて第 1 の機械学習モデルにより音響特徴量と感情状態との関連を取得して状態判断テーブルに学習結果を保存する感情状態学習ステップ、

音楽選択学習手段により、予め備えた音楽選択学習テーブルに基づいて第 2 の機械学習モデルにより選択する音楽と感情状態との関連を取得して音楽選択テーブルに学習結果を保存する音楽選択学習ステップ、

音響情報入力手段により、該音声及び該活動音を取得し、音響情報として入力する音響情報入力ステップ、

音響特徴量抽出手段により、該音響情報から該音響情報の音響スペクトルに係る特徴量を抽出する音響特徴量抽出ステップ、

状態判断手段により、該音響特徴量を入力して、該状態判断テーブルに基づいて音響情報に係る人間の感情状態を該第 1 の機械学習モデルにより判断する状態判断ステップ、

音楽選択手段により、該感情状態を入力して、該音楽選択テーブルに基づいて該第 2 の機械学習モデルにより音楽を選択する音楽選択ステップ、

音楽出力手段により、該音楽を背景音楽として出力する音楽出力ステップ

の各ステップを含むことを特徴とする人間の感情状態に応じた音楽出力方法。

30

40

【請求項 7】

前記感情状態学習ステップにおいて、

前記感情状態学習用テーブルを、前記音声及び活動音を生じさせる人間の特徴別に用意して学習し、

前記状態判断ステップにおいて、

前記状態判断手段が、判断時に選択された特徴別の学習結果を用いて音響情報に係る当該人間の感情状態を判断する

請求項 6 に記載の人間の感情状態に応じた音楽出力方法。

【請求項 8】

前記状態判断テーブルにおいて、

50

同一の音響特徴量に対して、複数の感情状態の確度が定義される
請求項 6 又は 7 に記載の人間の感情状態に応じた音楽出力方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音楽を選択し出力する音楽出力装置及び音楽出力方法に係り、より詳しくは人間の発する音声や活動音或いはテキストから人間の感情状態を推定してその感情状態に適した音楽を選択し、出力する技術に係るものである。

【背景技術】

【0002】

音楽が人間の感情を高めたり抑制したりする効果をもつことが知られている。また演劇やドラマなどでは、登場人物の感情状態を表現し、あるいは視聴者の感情を盛り立てる目的で背景音楽が多用されている。

このような背景音楽の選択や出力は、従来熟練した音響効果の専門家に委ねられており、日常生活において手軽に利用できる技術ではない。

【0003】

例えば、特許文献 1 では、利用者が付与したいと思う意図する感情を手動により選択することで背景音楽信号を音声信号に重畳して再生する音声再生データ出力装置が提案されている。本技術は利用者が自発的に音楽を再生することを企図したものであり、従来専門家でなければ難かった感情に合わせた音楽選択を感情・音声データベースを用いることで、容易に選択ができるようにした点が優れている。

【0004】

【特許文献 1】特開 2001 - 166786 号公報

【0005】

また、音声情報から話者の感情状態を認識する方法が知られている。例えば特許文献 2 に開示される対話処理装置では、ユーザから入力された語句の概念を抽出し、その語句の概念に基づいて、ユーザの感情を推定し、その感情を表す感情情報を出力する。さらに、感情情報に基づいてユーザに出力する出力文を生成する機能を有している。

【0006】

【特許文献 2】特開 2001 - 215993 号公報

【0007】

本技術は、音響情報を用いてこれを音声認識し、得られた語句の概念が抽出され、その概念に基づいて、ユーザの感情が推定される。そして、その結果推定された感情情報に基づいて、ユーザに出力する出力文が生成される。また、ロボットから合成音を出力してユーザに应答するなど、ユーザインタフェースとして提供することも提案されている。

【0008】

さらに、特許文献 3 に開示される音声制御装置では、話者の意図にあった細かな制御を目的として、人間の状態（感情、体調等）による音声指示の変化を的確に判断し、単なる音声命令だけでなく、その状態を加味して出力する装置を開示している。

本装置によって音声命令に含まれるニュアンスにも対応した制御を可能にしており、例えばテレビのリモコンに搭載して音量調節の程度を変化させることなどが可能である。

【0009】

【特許文献 3】特開平 10 - 55194 号公報

【0010】

特許文献 4 に開示される構成では、音声認識時に、感情の変化に伴う音韻スペクトルの変形を加味し、認識結果と共に音声もっている話者の感情の度合を示す感情レベルを出力する技術を開示している。

【0011】

【特許文献 4】特開平 11 - 119791 号公報

【0012】

10

20

30

40

50

上記は話者の音声から感情状態を取得する技術であるが、その他にテキストから人間の感情状態を取得する技術としては特許文献5に開示されるような技術もある。

本技術は、電子文書が有する感情を自動的に判定すると共に、その判定の精度を高くするものであり、単語の感情表現を感情ベクトル値で表した感情表現辞書を用いることを特徴としている。

【0013】

【特許文献5】特開2002-230011号公報

【発明の開示】

【発明が解決しようとする課題】

【0014】

上述した従来技術のように音声から人間の感情の状態を得る方法が開発されているが、特許文献2の方法は、ユーザとの対話を目的としてユーザの感情状態を知ろうとするものであり、特許文献3及び4はユーザの感情状態を知ることによってユーザが行った命令の程度を決定するものである。

また特許文献5では得られた感情を別のユーザに視覚的に知らせるものであって、聴覚的な効果を企図したものではない。

【0015】

本発明は、音響情報やテキスト情報から人間の感情状態を推定し、当該情報の発信者（話し手や書き手）や受信者（聞き手や読み手）の感情状態を効果的にコントロールできるように音楽を出力するものであり、上記従来技術のいずれによっても実現されていない。

【0016】

本発明は従来技術で提供できなかった、人間の感情状態に基づいて自動的に音楽出力すると共に、音楽選択の的確性、装置の簡易性を図ることを目的として創出したものである。

【課題を解決するための手段】

【0017】

本発明は、上記の課題を解決するために、次のような手段を用いる

すなわち、請求項1に係る発明は人間の発する音声及び活動に伴って生じる活動音に対して、当該人の感情状態に適し、該音声及び該活動音の背景音楽を選択して出力可能な音楽出力装置を提供する。

該装置において感情状態学習用テーブルと、該感情状態学習テーブルに基づいて第1の機械学習モデルにより音響特徴量と感情状態との関連を取得して状態判断テーブルに学習結果を保存する感情状態学習手段と、音楽選択学習用テーブルと、該音楽選択学習テーブルに基づいて第2の機械学習モデルにより選択する音楽と感情状態との関連を取得して音楽選択テーブルに学習結果を保存する音楽選択学習手段とを備える。

【0018】

さらに、該音声及び該活動音を取得し、音響情報として入力する音響情報入力手段と、該音響情報から該音響情報の音響スペクトルに係る特徴量を抽出する音響特徴量抽出手段と、該音響特徴量を入力して、該状態判断テーブルに基づいて音響情報に係る人間の感情状態を該第1の機械学習モデルにより判断する状態判断手段と、判断された該感情状態を入力して、該音楽選択テーブルに基づいて該第2の機械学習モデルにより音楽を選択する音楽選択手段と、該音楽を背景音楽として出力する音楽出力手段とを備える。

本構成において、音響スペクトルに係る特徴量として、強度、テンポ、及び周波数を全て用いることを特徴とする。

【0019】

ここで、前記感情状態学習用テーブルを、前記音声及び活動音を生じさせる人間の特徴別に用意して学習し、前記状態判断手段において、判断時に選択された特徴別の学習結果を用いて音響情報に係る当該人間の感情状態を判断する構成でもよい。

【0020】

上記の状態判断テーブルにおいて、同一の音響特徴量に対して、複数の感情状態の確度

10

20

30

40

50

が定義される構成でもよい。

【0021】

前記機械学習モデルが、最大エントロピーモデル、サポートベクトルマシンモデルのいずれかを用いる構成でもよい。

【0022】

前記音楽出力装置が、電話回線との接続手段を備え、音声情報入力手段が、電話回線における通話音声を取得して入力すると共に、音楽出力手段が、電話回線における通話音声に音楽を挿入する構成でもよい。

【0023】

本発明は、人間の発する音声及び活動に伴って生じる活動音に対して、当該人の感情状態に適し、該音声及び該活動音の背景音楽を選択して出力可能な音楽出力方法として提供することもできる。

10

該方法においては、まず、感情状態学習手段により、予め備えた感情状態学習テーブルに基づいて第1の機械学習モデルにより音響特徴量と感情状態との関連を取得して状態判断テーブルに学習結果を保存する感情状態学習ステップ、音楽選択学習手段により、予め備えた音楽選択学習テーブルに基づいて第2の機械学習モデルにより選択する音楽と感情状態との関連を取得して音楽選択テーブルに学習結果を保存する音楽選択学習ステップを有する。

【0024】

各学習ステップの後、音響情報入力手段により、該音声及び該活動音を取得し、音響情報として入力する音響情報入力ステップ、音響特徴量抽出手段により、該音響情報から該音響情報の音響スペクトルに係る特徴量を抽出する音響特徴量抽出ステップ、状態判断手段により、該音響特徴量を入力して、該状態判断テーブルに基づいて音響情報に係る人間の感情状態を該第1の機械学習モデルにより判断する状態判断ステップ、音楽選択手段により、該感情状態を入力して、該音楽選択テーブルに基づいて該第2の機械学習モデルにより音楽を選択する音楽選択ステップ、音楽出力手段により、該音楽を背景音楽として出力する音楽出力ステップの各ステップを含むことを特徴とする。

20

【0025】

ここで、感情状態学習ステップにおいて、感情状態学習用テーブルを、前記音声及び活動音を生じさせる人間の特徵別に用意して学習し、状態判断ステップにおいて、状態判断手段が、判断時に選択された特徴別の学習結果を用いて音響情報に係る当該人間の感情状態を判断する構成でもよい。

30

【0026】

前記状態判断テーブルにおいて、同一の音響特徴量に対して、複数の感情状態の確度が定義される構成でもよい。

【発明の効果】

【0027】

以上の発明により次の効果を奏する。すなわち、請求項1に記載の発明によると、人間の発する音声及び活動音から人間の感情状態を的確に判断し、自動的に音楽を出力できる音楽出力装置を提供することができる。

40

【0028】

本発明では、感情状態の判断の際と、音楽選択の際のそれぞれの段階で機械学習モデルを用いた処理を行う。

すなわち、感情状態の判断に際して機械学習を用いることができるので、ユーザの個性に応じた状態判断に寄与し、より正確な状態判断を行うことが可能である。

また、音楽選択に際して機械学習を用いることができるので、感情状態に応じた適切な音楽の選択に寄与する。すなわち、感情状態に応じて画一的な対応動作を取るのではなく、複数のパラメータで表される感情状態に対しても、最適な音楽選択に寄与する。

請求項6記載の発明についても同様の効果を奏する。

【0029】

50

請求項 2 に記載の発明によれば、全てのユーザについて学習させずとも、例えば「早口の男性」「声の低い女性」など、ある程度特徴を分けて学習させ、使用時に選択するだけでも判断の精度を上げることができる。

請求項 7 記載の発明についても同様の効果を奏する。

【0030】

請求項 3 に記載の発明によれば、従来の方法では 1 つの感情状態の程度でしか表すことができなかつたのに対し、本発明では複数の感情状態に対して同時にそれらの確度を決定することができる。

請求項 8 記載の発明についても同様の効果を奏する。

【0031】

請求項 4 に記載の発明では、機械学習の学習モデルとして公知の最大エントロピーモデル、サポートベクトルマシンモデルを用いることで、既知の学習エンジンを採用し、簡便に本発明を実施することができる。

【0032】

請求項 5 に記載の発明では、電話における通話時に感情状態に合わせた音楽の挿入を行うことができるので、快適な通話環境の提供に寄与する。

【発明を実施するための最良の形態】

【0033】

以下、本発明の実施形態を、図面に示す実施例を基に説明する。なお、実施形態は下記に限定されるものではない。

【実施例 1】

【0034】

以下、本発明の実施形態を、図面に示す実施例を基に説明する。なお、実施形態は下記に限定されるものではない。

図 1 は本発明の第 1 の実施形態に係る人間の感情状態に応じた音声出力装置 (1) の構成図である。本装置は、例えば公知の音声入力端子又はマイクを備えたパーソナルコンピュータによって実現することが簡便である。すなわちハードウェアとしては集音マイクなどを接続して人間の音声及び活動時の音を入力するための音響入力端子 (10) と、各種信号処理・演算処理を行う CPU (20)、該 CPU (20) に接続され必要に応じてデータの入出力を行う外部記憶装置であるハードディスク (30) の他、選択された音楽をライン出力するための音楽出力端子 (40) や、直接出力するスピーカ (41) を備えている。

【0035】

そして、CPU (20) の音響入力部 (21) では、集音マイクにより取得した音声・活動音をコンピュータで処理可能なデジタル信号の音響情報に変換処理する。一般にアナログデジタル変換を行うものであり、特にサウンドボードによって処理される場合が多い。

【0036】

次に、音響特徴量抽出部において、入力された音響情報から音響スペクトルを取得し、その音響特徴量を抽出する。前述の通り音響情報には音声の他、活動音も含まれており、いずれであっても感情状態を判断することができる。

音声の場合には、非特許文献 6 に開示されるような公知のソフトウェアを実行して音声情報から音声波形・スペクトルを算出処理し、音声の特徴量を抽出するのが簡便である。

【0037】

【非特許文献 6】Windows (登録商標) 対応音声処理ソフトウェア 「音声工房 Pro」<http://www.sp4win.com/sp4win/pro/pro.htm>

【0038】

本発明では、例えばユーザが定期的に発声する単語を入力して、その際のユーザの感情状態を判断するように構成し、例えば「オーライ」という単語についてユーザが「元気」

10

20

30

40

50

なのか、「眠い」のか、「怒っている」のか、といった感情状態の判断を行う。

【 0 0 3 9 】

図 2 は音響特徴量抽出部 (2 2) 及び状態判断部 (2 3) における処理を詳細に説明する流れ図である。図のように、音響特徴量抽出部 (2 2) では、音韻スペクトルから強度 (5 1)、テンポ (5 2)、周波数 (5 3) をそれぞれ測定する。

具体的には上述した公知のソフトウェア等により、図 3 に示す音韻スペクトルが得られる。図 3 において上段は、感情状態が「元気」(通常)であるときの音声波形、下段は該音声の強度(パワー)を示しており、概ね 1 . 3 秒程度の発声時間内の時間的变化である。

さらに、図 4 は感情状態が「眠い」時の音声波形(上段)及び強度(下段)を示している。

10

【 0 0 4 0 】

各感情状態における強度を比較すると、「元気」時の最大パワーが - 5 d B 程度、「眠い」時の最大パワーが - 1 0 d B 程度であり、「元気」時には大きな発声が行われていることがわかる。強度測定 (5 1) では、最大強度の他、最低強度、平均強度、最低強度と最大強度の差などを測定値とすることができる。

【 0 0 4 1 】

音響特徴量抽出部 (2 2) においてテンポを抽出する際には、公知の任意の手法を用いることができるが、例えば強度が所定の閾値を超えた各時刻を取得し、その時刻の間の時間をテンポとして得ることが考えられる。

20

すわなち、図 3 において、最初に閾値を超えた時刻を t_1 、再び閾値以下になった後に再度閾値を超えた時刻を t_2 とすると、これらはそれぞれ「オー」の発声開始時刻、「ライ」の発声開始時刻であり、この間の時間 $t_2 - t_1$ がテンポとなる。

【 0 0 4 2 】

一般的にユーザが「眠い」感情状態の時、話すテンポが遅くなることが知られており、図 4 に着目すると、上記の時刻は音声波形上でそれぞれ t_3 、 t_4 のときであることがわかる。図 3、図 4 の横座標はそれぞれ同一であり、両図の比較から明らかにテンポが遅くなっている。

テンポ測定 (5 2) では、上記のように各音素の開始時刻間の間隔や、各音素における最大強度となった時刻間の間隔、あるいは強度のグラフから各音素間で強度が 0 に近づき谷となる間の時間を求めることもできる。

30

【 0 0 4 3 】

次に、図 5 及び図 6 は音声情報からピッチ(基本周波数)の時間変化を示したものである。両図の比較においても、基本周波数の最低値及び最大値が、「元気」時にはそれぞれ 1 2 0 H z、1 4 0 H z であるのに対し、「眠い」時には 1 0 0 H z、1 1 0 H z 程度とやや低くなることがわかる。

周波数測定 (5 3) では、ピッチの最大値および最小値、又は平均値などを測定値として用いることができる。

【 0 0 4 4 】

本発明の音響特徴量抽出部 (2 2) ではこのように基本周波数を取得する構成の他、音声波形のスペクトルの中で、特にエネルギーが集中している周波数成分で、声道の共鳴周波数を表すホルマントを抽出する構成をとることもできる。

40

図 7 及び図 8 はホルマント周波数を示すグラフである。両図の比較から「元気」時にはホルマント周波数の分散が大きいのに対し、「眠い」時には平坦な分布となっており、全体に抑揚が少なくなっている。

この場合、各分布値から分散を算出し、比較することにより、感情状態の特性を得ることができる。

【 0 0 4 5 】

音響特徴量抽出部 (2 2) では以上のように音響スペクトルから強度、テンポ、周波数の各特徴量を得て、状態判断部 (2 3) に送る。これらの各特徴量は C P U (2 0) に接

50

続される公知のメモリ（図示しない）等に一時的に記録されるか、ハードディスク（30）に格納される。

【0046】

また、活動音の場合についても、音声と同様に音響情報から音響スペクトルを取得し、強度やテンポ、周波数について測定する。ここで活動音とは人間の活動に伴って生じる音全てを指し、例えば調理場においては野菜を切る音、肉を焼く音、食器を洗う音などが該当する。その他にも日常的に掃除機をかける音や、階段を昇降する音など、作業時に特徴的な音が生じる場面は多い。

【0047】

そして、例えば階段を昇降する音の場合に、そのテンポが短く音が大きい場合にはユーザが「慌てている」と判断できるし、テンポが不規則で強弱の変化が激しい場合には「注意力が散漫になっている」、軽快に適正なテンポで昇降している場合には「快調」と判断することもできる。

さらに、周波数によって昇降する人の分類とその気分の判断を行うことも可能であり、活動音により音声と同様に様々な感情状態の判断を行うことができる。

【0048】

さらに、活動音によっては、変化する感情状態をその時々で把握することは必ずしも容易でない場合がある。しかし、例えば掃除機をかけているときには気分でどのような音楽をユーザが聞きたいのか、あるいは食器を洗っているときにはどうか、といったようにその場面での平均的なユーザの感情を推定することが可能である。

すなわち、本発明におけるユーザの感情状態とは、時々刻々変化するユーザの感情状態だけでなく、一般的な生活場面での感情といったものまでを含む概念である。

【0049】

抽出された音響特徴量は状態判断部（23）において状態判断テーブル（31）と比較されて感情状態の判断に用いられる。

すなわち、本発明において感情状態を判断する際には複数の音響特徴量から、状態判断テーブル（31）との近似度を点数化し、合計点の最も高いものをユーザの感情状態と判断する。

【0050】

例えば、ユーザの感情状態として、感情状態A、感情状態B、感情状態Cの3つが予め分類されており、それぞれの確度を順次算出する。これらの3つは上記の「元気」「眠い」「怒っている」などに対応している。

そして、まず強度測定（51）の結果から、感情状態判断テーブル（31）を参照して強度から見たときの確度を決定する。次の表1は感情状態判断テーブル（31）の一例である。

【0051】

【表1】

最大強度 (dB)	ユーザの状態	確度
0未満～-2以上	怒っている	0.8
-2未満～-5以上	怒っている	0.5
-5未満	怒っている	0.2
0未満～-8以上	元気	0.7
-8未満～-15以上	眠い	0.5
-15未満	眠い	0.8

【0052】

本発明では、強度に対して複数のユーザ状態が重複して与えられていることを特徴とし、例えば最大強度が0～-2dBであった場合には、「怒っている」確度は0.8、「元気」である確度は0.7、「眠い」の確度は与えられていないため確度を0として設定する。

10

20

30

40

50

すなわち、強度に関しては各感情状態確度算出処理(54)(55)(56)でそれぞれの感情状態の確度が上記感情状態判断テーブル(31)から決定される。

【0053】

次に、テンポ測定(52)の結果についても感情状態判断テーブル(31)に備えられた次のような表から各感情状態の確度を決定する。

例えば、図3、図4の音響スペクトルでは、「元気」時のテンポが0.22秒、「眠い」時のテンポが0.55秒程度であり、感情状態判断テーブル(31)からテンポに関しては、図3の音声情報が入力されると、「怒っている」確度が0.3、「元気」確度が0.8、「眠い」確度が0.1となる。

【0054】

【表2】

音素間の間隔 (s)	ユーザの状態	確度
～0.2未満	怒っている	0.9
0.2以上～0.5未満	怒っている	0.3
～0.2未満	元気	0.6
0.2以上～0.5未満	元気	0.8
～0.4未満	眠い	0.1
0.4以上～	眠い	0.9

【0055】

同様に、周波数測定(53)の結果についても感情状態判断テーブル(31)に備えられた表3から各感情状態の確度を決定する。

図5、図6の音響スペクトルによれば、「元気」時の基本周波数の最大値及び最小値の中間値(平均値)は130Hzであり、この場合に「怒っている」確度は0.6、「元気」な確度は0.7、「眠い」確度は0となる。

【0056】

【表3】

平均周波数 (Hz)	ユーザの状態	確度
130以上	怒っている	0.6
110以上～130未満	怒っている	0.3
120以上	元気	0.7
120未満	元気	0.2
120未満	眠い	0.8

【0057】

従って、感情状態A確度算出処理(54)で例えば「元気」な確度を算出する場合には、音声情報として「元気」時の音韻スペクトルが得られたときに強度からは0.7、テンポからは0.8、周波数からは0.7が決定し、各合計は2.2である。このとき、ユーザの感情状態が「元気」な確度は2.2と算出される。

【0058】

さらに感情状態B確度算出処理(55)で「怒っている」確度を算出する場合、同様に「元気」時の音韻スペクトルからは、強度に関して0.5、テンポに関して0.3、周波数に関して0.6が得られるから、「怒っている」確度は1.4と算出される。

同様に感情状態C確度算出処理(56)で「眠い」確度は、強度に関して0、テンポに関して0.1、周波数に関して0であり、確度0.1と算出される。

【0059】

以上の結果、状態判断部(23)では図3のような音声情報が入力された場合には、「元気」な確度が2.2、「怒っている」確度が1.4、「眠い」確度が0.1であるとユーザの感情状態を判断する。

このように、従来の方法では1つの感情状態の程度でしか表すことができなかったのに

10

20

30

40

50

対し、本発明では複数の感情状態に対して同時にそれらの確度を決定することができる。

【0060】

また、活動音の場合にも同様の方法によりユーザの感情状態を判断できる他、音響情報から「洗濯時の気分」「食事を用意する際の気分」といった場面毎の感情状態について、そのいずれであるかを状態判断部(23)で判断するように構成することもできる。

【0061】

次に音楽選択部(24)の動作を説明する。図9には対応動作選択処理の流れ図を示す。

状態判断部(23)で判断された感情状態から、音楽選択テーブル(32)を参照し、出力すべき音楽の種類を選択する。音楽の種類としては音楽ジャンルによる分類、楽曲の印象尺度による分類など任意の分類方法によることができる。

楽曲印象の分類としては、本件出願人らによる楽曲の分類尺度を用いることもできる。

【0062】

【特許文献7】特開2004-118010号公報

【特許文献8】特開2004-118011号公報

【0063】

ここでは説明に便利のように音楽の種類を「明るい曲」「静かな曲」「勇ましい曲」の3種類とし、どの種類に属するかを順次判定する。

音楽選択テーブル(32)には感情状態と、出力すべき音楽との対照表が格納されている。表4は対照表の一例である。

【0064】

【表4】

感情状態A確度	感情状態B確度	感情状態C確度	出力すべき音楽
2.2	1.4	0.1	「明るい曲」
1.5	2.1	0.3	「静かな曲」
0.5	0.2	1.9	「勇ましい曲」

【0065】

状態判断結果を音楽選択部(24)に入力すると、音楽選択テーブル(32)の音楽の種類、「明るい曲」「静かな曲」「勇ましい曲」をそれぞれ出力するか(61)(62)(63)順次判定する。

判定は、択一的にいずれの音楽にするか一度にしてもよいし、図9のように順次実行してもよい。順次実行する場合には、上位の判定で対応動作を実行した場合には下位の判定を行わない構成でもよい。

【0066】

判定には公知の手法により上位対照表との類似度を調べることが簡便であり、例えば最も簡単には各感情状態の確度と、状態判断部(23)から得られた各感情状態の確度とのそれぞれの差の絶対値を求め、その値が所定の閾値以下となる場合に実行すべき対応動作を実行する判定を行えばよい。

【0067】

一例として、感情状態A、B、Cの確度がそれぞれ2.0、1.8、0.3と入力されたとき、確度の差の絶対値は、表4の「明るい曲」の各確度より、0.2、0.4、0.2であるから合計は0.8になる。

ここで例えば閾値が1.0であった場合、「明るい曲」を出力すべきと判定(61)する。

【0068】

選択された結果(64)は、音楽出力部において楽曲データ(33)から該当する楽曲情報を読み出し、出力端子(40)から出力又はスピーカ(41)により再生される。

楽曲データ(33)は表5のように構成されており、各楽曲のファイルとその楽曲の種類が対応付けされて格納される。データは、MIDIファイルやMP3による。

10

20

30

40

50

そして、音楽出力部（25）では公知の音楽再生処理方法により、再生を行う。

【0069】

【表5】

種類	楽曲情報
「明るい曲」	ピース.mp3
「明るい曲」	勝手にシンドバッド.mp3
「静かな曲」	ます.mp3
「勇ましい曲」	トルコ行進曲.mp3

10

【0070】

以上の構成により、本発明はマイクなどで取得した音声・活動音を入力し、音響特徴量抽出部（22）によって特徴量が抽出され、状態判断部（23）において「眠い」と判断された場合に、音楽選択部（24）で「勇ましい曲」と選択される。その結果、音楽出力部（25）からは「勇ましい曲」に対応するトルコ行進曲.mp3の楽曲情報が再生処理されることになる。

【0071】

同様に、活動音として階段を慌ただしく駆け上がっている時には、ゆっくりしたテンポの曲を流して落ち着いて昇降するように促したり、家事を行っている場合にはリズムカルな曲を流して素早い作業を助けたり、といった使用に寄与する。

20

従来の手動による音楽選択と異なり、予め音楽選択テーブルに蓄えた情報に従って自動的に最適な音楽を選択し出力することができる。

【実施例2】

【0072】

図10には、本発明の第2の実施形態による音楽出力装置（1a）の構成図を示す。本実施例は上記第1の実施形態に加えて感情状態学習部（100）をCPU（20）に、該感情状態学習部（100）で学習に用いるための学習データが格納された感情状態学習テーブル（101）をハードディスク（30）に備えている。

【0073】

感情状態学習テーブル（101）には、ユーザが様々な感情状態にあるときの音声と、その際の感情状態を主観的及び客観的に評価した結果が格納されており、例えば表6のように構成されている。

30

【0074】

【表6】

ユーザの音声	感情状態
音声波形1	感情状態A
音声波形2	感情状態A
音声波形3	感情状態B
音声波形4	感情状態C
音声波形5	感情状態C
音声波形6	感情状態C

40

【0075】

そして、感情状態学習テーブル（101）はまず音響特徴量抽出部（22）に入力されて、実施例1の時と同様に音響特徴量を抽出する。抽出された音響特徴量は、感情状態学習部（100）に入力し、本実施形態に係る学習処理を行う。

ここで感情状態の学習には公知の機械学習モデルを用いることができるが、一例としてサポートベクトルマシンを用いる方法を次に説述する。

【0076】

サポートベクトルマシン（以下、SVM）については、パターン認識などにおいて複数

50

の特徴量（特徴ベクトルで表される）から、その対象がどのクラスに属するかを判定する識別器として用いられている。

本件発明者らによる論文（非特許文献9）では、これを自然言語処理に適用し、SVMを用いた単語多義性解消問題に対する解法を示している。

【0077】

【非特許文献9】「SENSEVAL2J辞書タスクでのCRLの取り組み」電子情報通信学会 NL C2001-40

【0078】

SVMは分類の数が2個のデータを扱うものであるため、本論文ではペアワイズ法と呼ばれる手法を組み合わせ、単語多義性解消に必要な3個以上の分類を可能にする構成を開示している。

10

SVMは、空間を超平面で分割することにより2つの分類からなるデータを分類する方法である。このとき、2つの分類が正例と負例とからなるものとする、学習データにおける正例と負例の間隔（マージン）が大きいものほどオープンデータで誤った分類をする可能性が低いと考えられ、このマージンを最大にする超平面を求め、それを用いて分類を行う。

【0079】

図11はこのときのより小さなマージンの空間と、より大きなマージンの空間を表す図である。

図中において、白点(110)は正例、黒点(111)は負例を表し、実線は空間を分割する超平面(112)を、波線はマージン領域の境界を表す面(113)を意味している。

20

【0080】

通常、学習データにおいてマージンの内部領域に少数の事例が含まれていてもよいとする手法の拡張や、超平面の線形の部分を非線形に拡張（カーネル関数の導入）がなされたものが用いられる。以下、カーネル関数を導入する手法について説明を続ける。

【0081】

この拡張された方法は、以下の識別関数を用いて分類することと等価であり、その識別関数の出力値が正か負かによって2つの分類を判別することができる。

SVMによる分類に関する定式化の一例について説明する。尚、ここで示す定式化は公知のものである。まず、 x を2次元ベクトル、 x_i を*i*番目のサポートベクトル、 K をカーネル関数、 y_i は、後述するようにサポートベクトル x_i ($i = 1, 2, \dots, l$, $y_i \in \{1, -1\}$)に対する期待される出力値である。 b はパラメータである。

30

【0082】

【数1】

$$f(x) = \operatorname{sgn} \left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \right)$$

$$b = -\frac{\max_{i, y_i=-1} b_i + \min_{i, y_i=1} b_i}{2}$$

40

$$b_i = \sum_{j=1}^l \alpha_j y_j K(x_j, x_i)$$

【0083】

関数sgnは、数2によって定義される。

【0084】

【数 2】

$$\text{sgn}(x) = \begin{cases} 1 & (x \geq 0) \\ -1 & (\text{otherwise}) \end{cases}$$

【0085】

また、各 α_i は数 4 と数 5 の制約のもと、数 3 の $L(\alpha)$ を最大にする場合のものである。

【0086】

【数 3】

10

$$L(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

【数 4】

$$0 \leq \alpha_i \leq C \quad (i = 1, \dots, l)$$

【数 5】

$$\sum_{i=1}^l \alpha_i y_i = 0$$

20

【0087】

また、カーネル関数 K は数 6 のものを用いている。

【0088】

【数 6】

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^d$$

【0089】

ここで、 C 、 d は実験的に設定される定数である。例えば C を 1、 d を 2 に固定することができるが、適宜設定することができる。

30

ここで、 $\alpha_i > 0$ となるサポートベクトル \mathbf{x}_i について、数 1 の和をとっている部分は、この事例のみを用いて計算される。つまり実際の解析には学習データのうち、サポートベクトルと呼ばれる事例のみしか用いられない。

【0090】

以上に示すように、サポートベクトルマシン法は、分類の数が 2 個のデータを扱うもので、より多くの分類を行う場合に、従来では前記ペアワイズ手法など別の手法を組み合わせていた。

ところで、カーネル関数としては、ポリノミアル (Polynomial)、ガウシアン・ラジアル・ベシス・ファンクション (Gaussian Radial Basis Function)、エクスポネンシャル・ラジアル・ベシス・ファンクション (Exponential Radial Basis Function)、マルチレイヤー・パーセプション (Multi-Layer Perception)、フーリエ・シリーズ (Fourier Series)、スプライン (Splines)、ピースプライン (Bsplines)、アディティブ・カーネル (Additive Kernels)、テンソル・プロダクト・カーネル (Tensor Product Kernels) 等の関数を用いることもできる。

40

【0091】

上記の定式に従って、複数の特徴量を特徴ベクトル \mathbf{x}_i で与えたときに、サポートベクトル \mathbf{x}_i ($i = 1, 2, \dots, l$ 、 $y_i \in \{1, -1\}$) に対する期待される出力値 y_i を得

50

る。

感情状態学習部(100)においては感情状態学習テーブル(101)を用いてカーネル関数を定義し、状態判断テーブルに該関数を保存する。従って、本実施例では状態判断テーブル(31)は実施形態1で示した対応表ではなく、関数として格納されることになる。

【0092】

これに伴って、ユーザの感情状態を判断する場合には、まず音声入力部(21)から音声を入力して音響特徴量抽出部(22)で特徴量を抽出し、状態判断部(23)において学習モデルを用い、状態判断テーブル(31)を参照しながら、状態の判断結果を出力する。

10

本発明はこのような感情状態の判断に機械学習を用いた点に特徴を有し、機械学習モデルとしては上述したSVMの他、ニューラルネットワーク手法、最大エントロピー手法などを適宜用いることができる。

【0093】

本発明はこのような感情状態の判断に機械学習を用いた点に特徴を有し、機械学習モデルとしては上述したSVMの他、ニューラルネットワーク手法、最大エントロピー手法などを適宜用いることができる。

ニューラルネットワーク手法の場合には、入力に対応する望ましい出力の組を例示してニューラルネットワークの重みを決定する方法であるバックプロパゲーション(逆引き伝搬)を用いるのが好ましく、非特許文献10、11などに開示されている。

20

【0094】

【非特許文献10】「パターン識別」 新技術コミュニケーションズ、Richard O. Duda 著、尾上守夫訳、

【非特許文献11】<http://sail.i.ishikawa-nct.ac.jp/pattern/nn/nn3.html>

【0095】

最大エントロピー手法の場合にも、例えば非特許文献12に開示される本件発明者らによる手法によれば、複数の特徴量(特徴ベクトルで表される)から、その対象がどのクラスに属するかを判定する識別器として機能させることができる。

【0096】

【非特許文献12】「機械学習手法を用いた名詞句の指示性の推定」 自然言語処理、村田 真樹, 内元 清貴, 馬 青, 井佐原 均、2000年 1月, 7巻, 1号, p.31 ~ p.50

30

【実施例3】

【0097】

図12には、本発明の第3の実施形態による音楽出力装置(1c)の構成図を示す。本実施例は上記第1の実施形態に加えて音楽選択学習部(120)をCPU(20)に、該音楽選択学習部(120)で学習に用いるための学習データが格納された音楽選択学習テーブル(121)をハードディスク(30)に備えている。

【0098】

音楽選択学習テーブル(121)には、状態判断部(23)による判断結果のサンプルと、それに対する適切な対応動作を格納している。一例としては76のような態様である。なお、ここで状態判断結果は実施形態1で示したように、いくつかの次元を有するベクトル量として表すことができるので、表中でもベクトル量として定義している。

40

【0099】

【表 7】

状態判断結果	出力すべき音楽
状態判断ベクトル 1	「明るい曲」
状態判断ベクトル 2	「明るい曲」
状態判断ベクトル 3	「静かな曲」
状態判断ベクトル 4	「勇ましい曲」
状態判断ベクトル 5	「勇ましい曲」
状態判断ベクトル 6	「悲しい曲」

【0100】

10

該音楽選択学習部(120)では音楽選択学習テーブル(121)により上記実施形態2と同様の方法で例えばSVMを用いて状態判断結果と、それに対応する最も適切な対応動作を選択できるように機械学習する。

その結果は関数の形式で、音楽選択テーブル(32)に記録される。学習結果は音楽選択部(24)で機械学習モデルを用いた選択処理に使用される。

【0101】

以上、実施形態2及び実施形態3は互いに組み合わせて用いることもできる。このような機械学習モデルを用いた方法によれば、従来のように画一的に感情状態を判断するのではなく、複数の音響スペクトルの特徴量を用いて、最も適当な感情状態の判断を行うことができる。また、ユーザに合わせて学習させることで、ユーザの話し方、音声の特徴などに係わらず、正確な感情状態の判断を行うことができる。

20

全てのユーザについて学習させずとも、例えば「早口の男性」「声の低い女性」など、ある程度特徴を分けて学習させ、使用時に選択するだけでも判断の精度を上げることができる。

したがって、本実施形態2、3は判定の高精度化に極めて寄与する構成である。

【実施例4】

【0102】

本発明の音声出力装置は、電話回線において用いることができる。具体的には上記入力端子(10)及び出力端子(40)を回線に接続し、電話回線上を流れる音声信号を取得すると共に、該音声から感情状態を判断して適当な音楽を出力する。このとき、該音声に重畳して出力することにより会話の背景音楽として音楽を流すことができる。

30

これにより、発話者の感情に合わせて情緒豊かな会話を楽しむことができるようになる。

【実施例5】

【0103】

本発明の第5の実施形態は、上記のように音響情報を入力するのではなく、テキスト情報を入力し、該テキストに最適な音楽を出力する装置(1c)である。

すなわち、図13に示すように公知の文書ファイルであるテキスト情報(130)を外部記憶装置からの読み込み、或いはネットワークにおける受信によってテキスト入力部(130)で取得する。

40

【0104】

そして、テキスト情報(130)から感情表現に係る語句を抽出し、前述の特徴量と同じように該語句について感情状態の確度を算出する。例えば次のような処理である。テキストとして「試験に合格した。来年からは楽しい学校生活が始まる。でも友達と離ればなれになるのは寂しい。」というテキスト情報(130)を入力する。

語句抽出部(131)は、まず表8に示すような状態判断テーブル(131)を参照し、その中の「含まれる単語」に属する語句を抽出する。

【0105】

【表 8】

含まれる単語	表現される感情	確度
「合格」	うれしい	0.6
「楽しい」	うれしい	0.5
「許せない」	怒っている	0.9
「がっかり」	残念に思う	0.8
「寂しい」	残念に思う	0.7

【0106】

そして、状態判断部(23)では、表現される感情毎に確度を算出し、上記のテキスト例では、「合格」と「楽しい」が含まれるため「うれしい」感情の確度が1.1、「怒っている」確度は0、「残念に思う」確度は0.7となる。

そして、音楽選択部(24)ではこれらの確度から、前記の方法と同様に音楽選択テーブル(32)を用いて最適な音楽を選択する。

【0107】

本実施形態においても感情状態学習部(100)や音楽選択学習部(120)を設けることができ、図13には後者を設けた例を示している。この場合、音楽選択学習テーブル(121)には、表現される感情の組と、選択すべき音楽の種類を複数用意し、学習に用いる。

【実施例6】

【0108】

実施例5の変形例として、図14に示すように入力端子(10)から人間の音声を音響情報として入力し、公知の音声認識処理を用いてテキスト化する構成(1d)をとることもできる。このため、音声入力部(140)と共に、音声認識処理部(141)を備える。

音声認識処理部(141)の構成については、特許文献2、4などに開示される他、市販の音声認識ソフトウェアなどを組み込んで採用することもできる。

認識後は、テキスト情報となるため、実施例5における構成と同様である。

【0109】

以上、本発明の実施形態として実施例1ないし6の構成を詳述した。本発明によれば、入力した音響情報から自動的に音楽を出力することができるので、実施例4における電話の背景音楽に限らず、様々な用途に用いられる。

例えば、演劇の公演において自動的に効果音を挿入することも可能である。舞台の音響をマイクにより集音して入力し、登場人物の感情状態に応じた効果音楽を自動的に再生することで、効果的な演出が可能となる。

また、会議室などに設定して、怒った口調を検出した場合には場を鎮めるような音楽を流す設備や、図書館内に設置して高揚した口調などふさわしくない音声を検出した場合に警告のチャイムを鳴らす、といった実施方法が考えられる。

【図面の簡単な説明】

【0110】

【図1】本発明の第1の実施形態に係る音声出力装置の構成図である。

【図2】本発明における感情状態を判断する流れ図である。

【図3】感情状態が通常時の音響スペクトルの強度を示すグラフである。

【図4】感情状態が眠い時の音響スペクトルの強度を示すグラフである。

【図5】感情状態が通常時の音響スペクトルの周波数を示すグラフである。

【図6】感情状態が眠い時の音響スペクトルの周波数を示すグラフである。

【図7】感情状態が通常時の音響スペクトルのホルマント周波数を示すグラフである。

【図8】感情状態が眠い時の音響スペクトルのホルマント周波数を示すグラフである。

【図9】本発明における音楽を選択する流れ図である。

10

20

30

40

50

- 【図10】本発明の第2の実施形態に係る音声出力装置の構成図である。
- 【図11】本発明に係るサポートベクトルマシンの説明図である。
- 【図12】本発明の第3の実施形態に係る音声出力装置の構成図である。
- 【図13】本発明の第5の実施形態に係る音声出力装置の構成図である。
- 【図14】本発明の第6の実施形態に係る音声出力装置の構成図である。

【符号の説明】

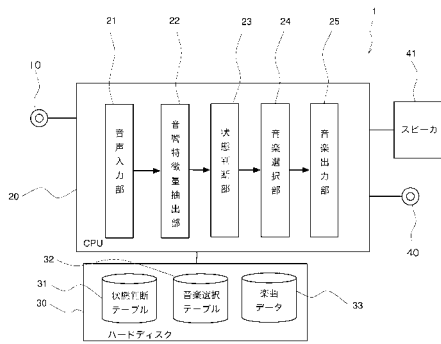
【0111】

- 1 人間の感情状態に応じた音声出力装置
- 10 音響入力端子
- 20 CPU
- 21 音声入力部
- 22 音響特徴量抽出部
- 23 状態判断部
- 24 音楽選択部
- 25 音楽出力部
- 30 ハードディスク
- 31 状態判断テーブル
- 32 音楽選択テーブル
- 40 音楽出力端子
- 41 スピーカ

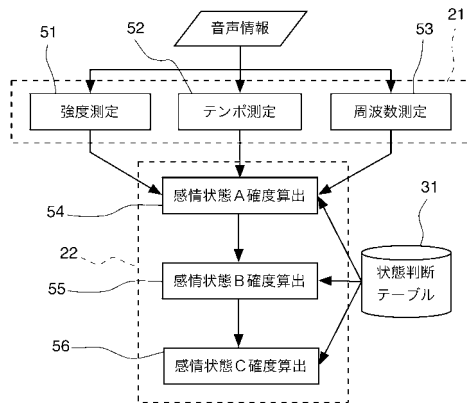
10

20

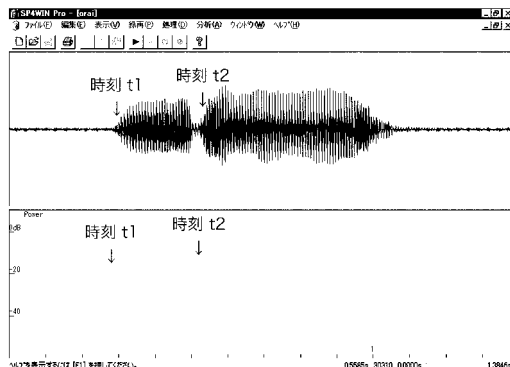
【図1】



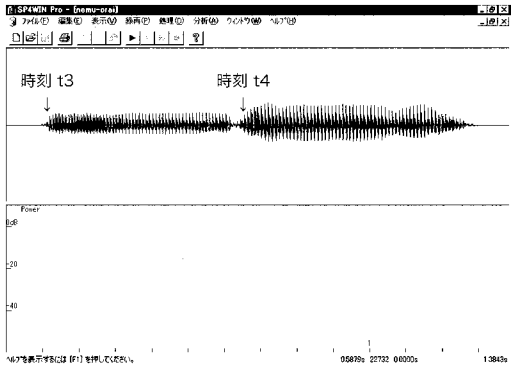
【図2】



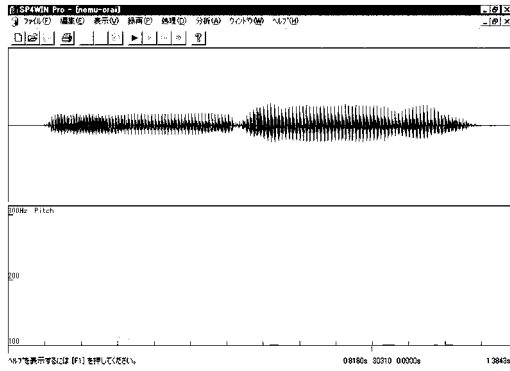
【図3】



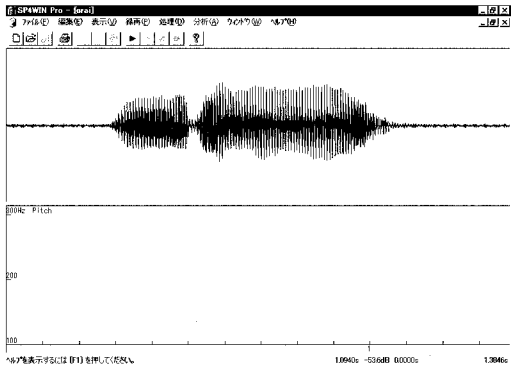
【図 4】



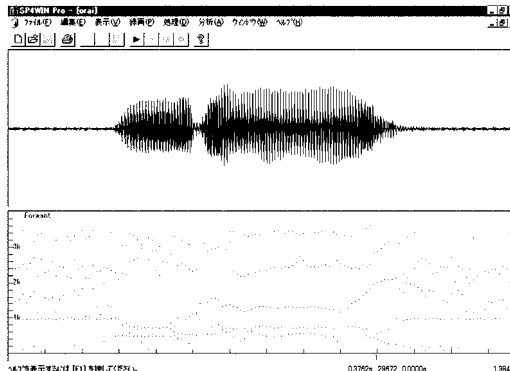
【図 6】



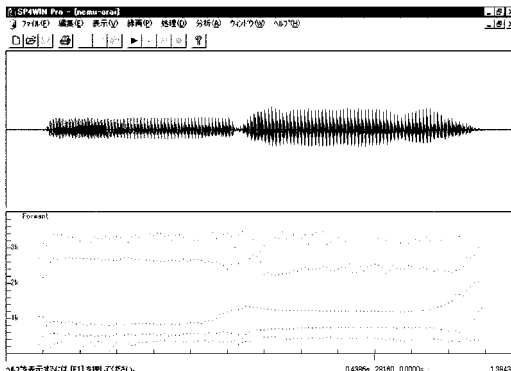
【図 5】



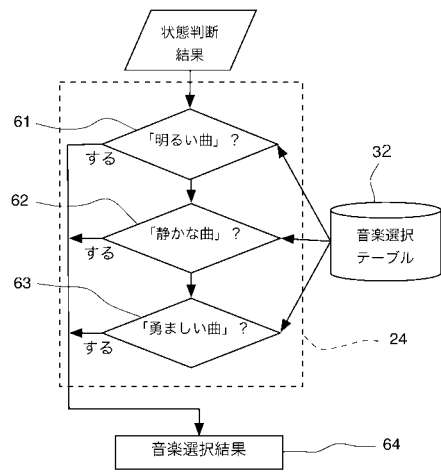
【図 7】



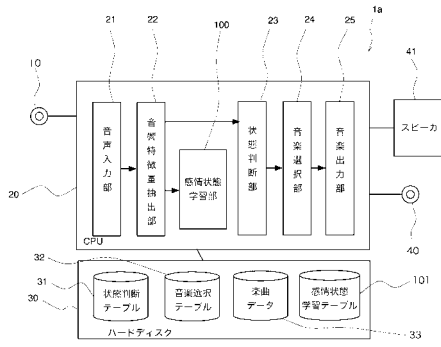
【図 8】



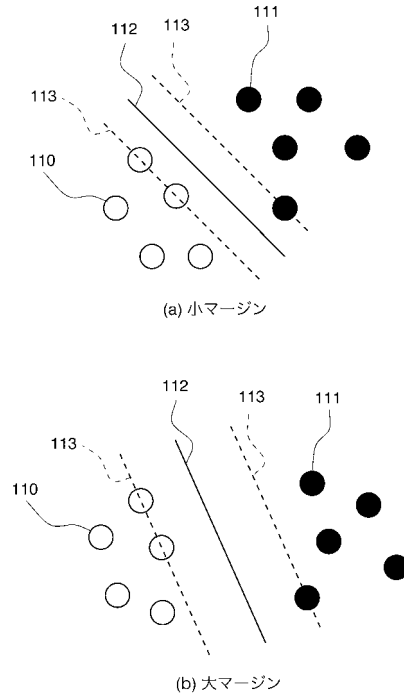
【図 9】



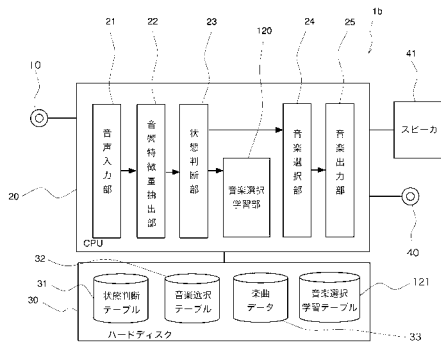
【図10】



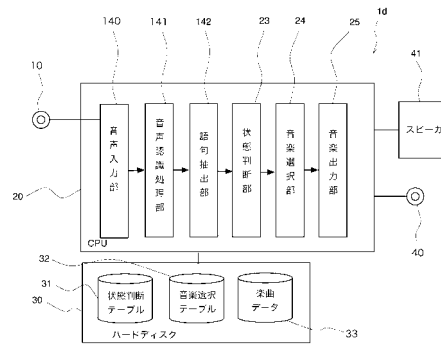
【図11】



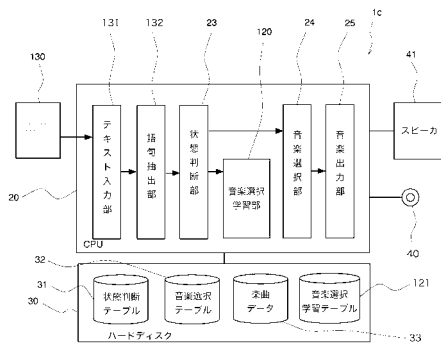
【図12】



【図14】



【図13】



フロントページの続き

- (56)参考文献 特開平10-214024(JP,A)
特開2003-132085(JP,A)
特開2002-032306(JP,A)
特開2003-210833(JP,A)
特開2003-345727(JP,A)
特開2000-268047(JP,A)
特開2001-162058(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 15/00-15/28
G10L 11/00-11/06
G10L 19/00-19/14