

(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2021-103202

(P2021-103202A)

(43) 公開日 令和3年7月15日(2021.7.15)

(51) Int. Cl.

G 1 O L 25/30 (2013.01)

F I

G 1 O L 25/30

テーマコード (参考)

審査請求 未請求 請求項の数 10 O L (全 27 頁)

(21) 出願番号 特願2019-233503 (P2019-233503)

(22) 出願日 令和1年12月24日 (2019.12.24)

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(74) 代理人 110002147

特許業務法人酒井国際特許事務所

(72) 発明者 松尾 直司

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(54) 【発明の名称】 更新プログラム、更新方法および情報処理装置

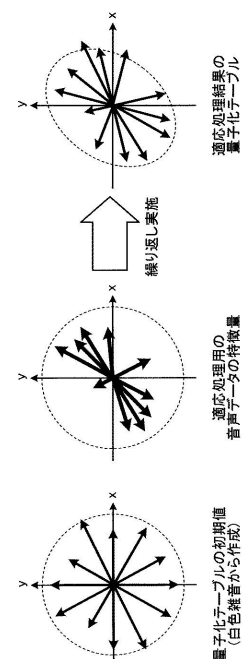
(57) 【要約】

【課題】 特定の会話状況を検出するためのキーワード設定を行わないで、特定の会話状況を検出することを課題とする。

【解決手段】 学習装置は、複数の量子化点を含む量子化テーブルについて、訓練データセットに含まれる複数の発話データの特徴量を量子化した量子化データに基づく、複数の量子化点それぞれの選択率を算出する。このようにして選択率を算出した学習装置は、複数の量子化点を選択率に基づき更新することで、量子化テーブルを更新する。

【選択図】 図7

量子化テーブルの適応制御の全体的な流れを説明する図



【特許請求の範囲】

【請求項 1】

コンピュータに、

複数の量子化点を含む量子化テーブルについて、訓練データセットに含まれる複数の発話データの特徴量を量子化した量子化データに基づく、前記複数の量子化点それぞれの選択率を算出し、

前記複数の量子化点を前記選択率に基づき更新することで、前記量子化テーブルを更新する、

処理を実行させることを特徴とする更新プログラム。

【請求項 2】

前記更新する処理は、前記複数の量子化点のうち、前記選択率が所定の基準以下の量子化点を前記量子化テーブルから除外し、除外した前記量子化点と異なる量子化点を前記量子化テーブルに追加し、前記選択率が所定の基準以下の量子化点以外の量子化点を選択結果に基づき更新する、ことを特徴とする請求項 1 に記載の更新プログラム。

【請求項 3】

前記算出する処理は、

前記複数の発話データそれぞれの前記特徴量に基づく各量子化データと、前記量子化テーブルに含まれる前記複数の量子化点それぞれとの距離とを算出し、

前記各量子化データの選択結果として、前記距離が最小となる量子化点を選択する、ことを特徴とする請求項 2 に記載の更新プログラム。

【請求項 4】

前記更新する処理は、前記選択率が所定の基準以下の量子化点を除外して、前記選択率が所定の基準以上である更新前の量子化点を新たに追加し、前記選択率が所定の基準以下の量子化点以外については、各量子化点を選択した量子化データの平均値に更新する、ことを特徴とする請求項 3 に記載の更新プログラム。

【請求項 5】

白色雑音に基づく複数の量子化点を初期値として含む前記量子化テーブルについて、第 1 の発話データから生成される各量子化データに基づく、前記複数の量子化点それぞれの選択率を算出し、

前記複数の量子化点のうち最も選択率が高い量子化点を無音相当の量子化点と特定する、処理を前記コンピュータに実行させ、

前記更新する処理は、前記無音相当の量子化点については、当該量子化点を選択した量子化データの平均値に更新し、前記無音相当以外の量子化点については前記選択率に基づく更新を実行する、ことを特徴とする請求項 4 に記載の更新プログラム。

【請求項 6】

前記更新する処理は、

前記無音相当の量子化点を除く更新後の量子化点に基づく量子化誤差を算出し、

前記量子化誤差が閾値以上の場合、第 2 の発話データを用いて、前記無音相当の量子化点の更新、および、前記無音相当以外の量子化点については前記選択率に基づく更新を実行し、

前記量子化誤差が閾値未満の場合、更新後の前記量子化テーブルを出力する、ことを特徴とする請求項 5 に記載の更新プログラム。

【請求項 7】

入力された音声情報と、更新された前記複数の量子化点を含む更新後の量子化テーブルとに基づくベクトル量子化により、前記音声情報の特徴量に該当する量子化点を対応付けた量子化結果を生成し、

前記量子化結果を、ニューラルネットワークを適用したモデルに入力した際に前記モデルから出力される出力情報が、前記量子化結果に対応する前記音声情報に所定の会話状況が含まれているか否かを示す正解情報に近づくように、前記モデルの学習を実行する、処理を前記コンピュータに実行させることを特徴とする請求項 1 に記載の更新プログラム。

10

20

30

40

50

【請求項 8】

判定対象の発話データの特徴量を量子化した量子化データを、前記学習を実行済みの前記モデルに入力することで取得された出力情報に基づいて、判定対象の発話データに前記所定の会話状況が含まれているか否かを判定する、

処理を前記コンピュータに実行させることを特徴とする請求項 7 に記載の更新プログラム。

【請求項 9】

コンピュータが、

複数の量子化点を含む量子化テーブルについて、訓練データセットに含まれる複数の発話データの特徴量を量子化した量子化データに基づく、前記複数の量子化点それぞれの選択率を算出し、

前記複数の量子化点を前記選択率に基づき更新することで、前記量子化テーブルを更新する、

処理を実行することを特徴とする更新方法。

【請求項 10】

複数の量子化点を含む量子化テーブルについて、訓練データセットに含まれる複数の発話データの特徴量を量子化した量子化データに基づく、前記複数の量子化点それぞれの選択率を算出する算出部と、

前記複数の量子化点を前記選択率に基づき更新することで、前記量子化テーブルを更新する更新部と、

を有することを特徴とする情報処理装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、更新プログラム、更新方法および情報処理装置に関する。

【背景技術】

【0002】

近年、コールセンターでは、オペレータと顧客との会話を録音し、録音した会話の情報を蓄積している。コールセンターには、蓄積された会話の情報を活用して、サービスを向上させたいというニーズがある。

【0003】

たとえば、蓄積された会話の情報を活用する技術として、次の様なものがある。顧客とオペレータとの会話に含まれる所定のキーワードの数を判定し、判定したキーワードの数に応じて、FAQ (Frequently Asked Questions) を表示したり、スーパーバイザへ通知したりする技術がある。また、オペレータの音声を文字列に変換し、文字列に伝達対象のキーワードが含まれているか否かをチェックすることで、オペレータが顧客に伝達内容を適切に伝えているかを判定する技術がある。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特開 2015-99304 号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

コールセンターでは、オペレータと顧客との会話において、予め指定した特定の会話状況が存在するか否かを検出したいというニーズがある。ここで、上記の技術を利用して、特定の会話状況が存在するか否かを検出する場合、検出したい会話状況に応じて、網羅的にキーワード設定を行い、設定したキーワードが、会話の情報に含まれているか否かを判定する処理を行うことが考えられる。

【0006】

しかしながら、特定の会話状況を検出するために、どれだけの数のキーワードを網羅すればよいかを事前に把握することは難しい。また、同じ意味の会話でも様々な言い回しがあるため、人手によってキーワードを網羅的に設定することは難しい。

【0007】

1つの側面では、本発明は、特定の会話状況を検出するためのキーワード設定を行わないで、特定の会話状況を検出することができる更新プログラム、更新方法および情報処理装置を提供することを目的とする。

【課題を解決するための手段】

【0008】

第1の案では、更新プログラムは、コンピュータに、複数の量子化点を含む量子化テーブルについて、訓練データセットに含まれる複数の発話データの特徴量を量子化した量子化データに基づく、前記複数の量子化点それぞれの選択率を算出する処理を実行させる。更新プログラムは、コンピュータに、前記複数の量子化点を前記選択率に基づき更新することで、前記量子化テーブルを更新する処理を実行させる。

【発明の効果】

【0009】

一実施形態によれば、特定の会話状況を検出するためのキーワード設定を行わないで、特定の会話状況を検出することができる。

【図面の簡単な説明】

【0010】

【図1】図1は、参考技術の一例を説明するための図である。

【図2】図2は、本実施例1に係る学習装置および判定装置の処理の一例を説明するための図である。

【図3】図3は、本実施例1に係るシステムの一例を示す図である。

【図4】図4は、本実施例1に係る学習装置の構成を示す機能ブロック図である。

【図5】図5は、本実施例1に係る生成部を説明するための図である。

【図6】図6は、量子化テーブルを生成する量子化テーブル生成部を説明する図である。

【図7】図7は、量子化テーブルの適応制御の全体的な流れを説明する図である。

【図8】図8は、量子化テーブルの初期値を説明する図である。

【図9】図9は、量子化点の選択を説明する図である。

【図10】図10は、量子化テーブルの更新を説明する図である。

【図11】図11は、量子化テーブルの更新前後を説明する図である。

【図12】図12は、量子化誤差による判定手法を説明する図である。

【図13】図13は、本実施例1に係るベクトル化部の処理を説明するための図である。

【図14】図14は、本実施例1に係るLSTMの一例を示す図である。

【図15】図15は、本実施例1に係るニューラルネットワークの一例を説明するための図である。

【図16】図16は、本実施例1に係る判定装置の構成を示す機能ブロック図である。

【図17】図17は、本実施例1にかかる量子化テーブルの生成処理の流れを示すフローチャートである。

【図18】図18は、本実施例1に係る学習装置の処理手順を示すフローチャートである。

【図19】図19は、本実施例1に係る判定装置の処理手順を示すフローチャートである。

【図20】図20は、異常な会話状況が含まれる音声データの内部ベクトルの一例を示す図である。

【図21】図21は、正常な会話の音声データの内部ベクトルの一例を示す図である。

【図22】図22は、第1計算部および第2計算部の拡張例を説明するための図である。

【図23】図23は、第3計算部のその他の処理を説明するための図である。

【図24】図24は、本実施例にかかるその他のシステムの一例を示す図である。

10

20

30

40

50

【図 2 5】図 2 5 は、ハードウェア構成例を説明する図である。

【発明を実施するための形態】

【0 0 1 1】

以下に、本願の開示する更新プログラム、更新方法および情報処理装置の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。また、各実施例は、矛盾のない範囲内で適宜組み合わせることができる。

【実施例 1】

【0 0 1 2】

[参考技術]

本実施例 1 に係る学習装置の処理を行う前に、参考技術について説明する。この参考技術は、従来技術ではない。図 1 は、参考技術の一例を説明するための図である。図 1 に示すように、参考技術は、音声認識部 5 A と、NN (Neural Network) 計算部 6 A とを有する。

10

【0 0 1 3】

音声認識部 5 A は、音声データが入力されると、音響モデル 5 B を基にして、音声データに含まれる文字列を抽出し、抽出した文字列が、キーワードリスト 5 C に定義されたキーワードに対応するか否かを認識する。音声認識部 5 A は、認識結果を、NN 計算部 6 A に出力する。

【0 0 1 4】

NN 計算部 6 A は、学習済みの NN モデル 6 C に基づく NN (図示略) に、音声認識部 5 A の認識結果を入力することで、音声データに特定の会話状況が含まれているか否かの出力値を算出する。ここで、特定の会話状況を適切に検出するためには、特定の会話状況に関するキーワードを網羅的にキーワードリスト 5 C に設定し、かかるキーワードリスト 5 C を用いて、NN モデル 6 C を学習することになる。しかし、特定の会話状況を検出できるようにするために、どれだけの数のキーワードを網羅すればよいかを、事前に把握することは難しい。また、異なるコールセンターに導入する際も、その都度、キーワードリスト 5 C をコールセンター固有のものに調整を行うことが望ましく、特定の会話状況を精度よく検出するためには、このキーワード設定および調整に多大な工数を要する。

20

【0 0 1 5】

[全体構成]

続いて、本実施例 1 に係る学習装置および判定装置の処理の一例について説明する。図 2 は、本実施例 1 に係る学習装置および判定装置の処理の一例を説明するための図である。図 2 に示すように、学習装置 1 0 0 は、学習用音声データベース 1 1 0 a と、生成部 1 2 0 と、第 1 計算部 1 3 0 と、第 3 計算部 1 4 0 と、第 2 計算部 1 5 0 と、学習部 1 6 0 とを有する。第 1 計算部 1 3 0、第 3 計算部 1 4 0、第 2 計算部 1 5 0、学習部 1 6 0 は、学習処理部に対応する。

30

【0 0 1 6】

学習用音声データベース 1 1 0 a は、各会話データを分割すること等により生成される複数の学習用の音声データを格納し、各学習用の音声データは、正解情報 1 1 0 b にそれぞれ対応付けられる。正解情報 1 1 0 b は、音声データごとに付与された、特定の会話状況が含まれるか否かを示す情報である。本実施例 1 では一例として、特定の会話状況を、「異常な会話状況」とする。異常な会話状況とは、顧客が不満を感じたり、怒り出したり、脅迫したりするなど、「通常でない状況」を含むものである。

40

【0 0 1 7】

生成部 1 2 0 は、学習用音声データベース 1 1 0 a から学習用の音声データを取得する。以下の学習装置 1 0 0 の説明において、学習用音声データベースから取得された学習用の音声データを、単に「音声データ」と表記する。生成部 1 2 0 は、音声データに対して、例えばベクトル量子化を行い、量子化結果の情報 (量子化系列) を生成する。たとえば、量子化系列は、発声の偏りを示す指標の一例である。生成部 1 2 0 は、各量子化結果を One Hot ベクトル化し、各量子化結果の One Hot ベクトルを、第 1 計算部 1 3 0 に出力す

50

る。

【0018】

第1計算部130は、再帰パスを持つ第1ネットワークに、量子化結果のOne Hotベクトルを順に入力し、第1ネットワークのパラメータに基づく計算を行うことで、内部ベクトルを算出する処理部である。たとえば、第1ネットワークは、LSTM (Long Short Term Memory) に対応する。第1計算部130は、音声データから生成される量子化結果のOne Hotベクトルを第1ネットワークに入力し、入力して得られる各内部ベクトルを、第3計算部140に出力する。

【0019】

第3計算部140は、第1計算部130から出力される複数の内部ベクトルを平均化する処理部である。第3計算部140は、平均化した内部ベクトルを、第2計算部150に出力する。以下の説明では、平均化した内部ベクトルを「平均ベクトル」を表記する。

【0020】

第2計算部150は、再帰パスを持たない第2ネットワークに、平均ベクトルを入力し、第2ネットワークのパラメータに基づく計算を行うことで、出力値（ニューロン値）を算出する処理部である。第2計算部150は、出力値を、学習部160に出力する。

【0021】

学習部160は、音声データを第1計算部130に入力した際に、第2計算部150から出力される出力値が、音声データに対応する正解情報110bに近づくように、第1計算部130のパラメータ、第2計算部150のパラメータを学習（誤差逆伝播法による学習）する。

【0022】

学習部160は、学習停止条件を満たすまで、誤差逆伝播学習を繰り返し実行し、LSTMモデル110c、DNN (Deep Neural Network) モデル110dを生成する。LSTMモデル110cは、学習済みの第1ネットワークのパラメータに対応する情報である。DNNモデル110dは、学習済みの第2ネットワークのパラメータに対応する情報である。学習装置100は、LSTMモデル110cの情報およびDNNモデル110dの情報を、判定装置200に通知する。なお、学習部160は、ネットワークを介して、LSTMモデル110cの情報およびDNNモデル110dの情報を、判定装置200に通知してもよいし、学習装置100と、判定装置200とを直接接続した上で、LSTMモデル110cの情報およびDNNモデル110dの情報を、判定装置200に通知してもよい。

【0023】

判定装置200は、生成部220と、第1計算部230と、第3計算部240と、第2計算部250と、判定部260とを有する。

【0024】

生成部220は、異常な会話状況であるか否かの検出対象となる音声データの入力を受け付ける。以下の判定装置200の説明において、異常な会話状況であるか否かの検出対象となる音声データを、単に、音声データと表記する。生成部220は、音声データに対して、例えばベクトル量子化を行い、量子化結果の情報を生成する。生成部220は、各量子化結果をOne Hotベクトル化し、各量子化結果のOne Hotベクトルを、第1計算部230に出力する。

【0025】

第1計算部230は、再帰パスを持つ第1ネットワークに、各量子化結果のOne Hotベクトルを順に入力し、第1ネットワークのパラメータに基づく計算を行うことで、内部ベクトルを算出する処理部である。第1計算部230は、第1ネットワークに設定するパラメータとして、LSTMモデル110cのパラメータを用いる。第1計算部230は、音声データから生成される量子化結果のOne Hotベクトルを第1ネットワークに入力し、入力して得られる各内部ベクトルを、第3計算部240に出力する。

【0026】

10

20

30

40

50

第3計算部240は、第1計算部130から出力される複数の内部ベクトルを平均化する処理部である。第3計算部140は、平均化した内部ベクトル（平均ベクトル）を、第2計算部250に出力する。

【0027】

第2計算部250は、再帰パスを持たない第2ネットワークに、平均ベクトルを入力し、第2ネットワークのパラメータに基づく計算を行うことで、出力値（ニューロン値）を算出する処理部である。第2計算部250は、第2ネットワークに設定するパラメータとして、DNNモデル110dのパラメータを用いる。第2計算部250は、出力値を、判定部260に出力する。

【0028】

判定部260は、第2計算部250から出力される出力値と、閾値とを比較して、音声データに、異常な会話状況が含まれているか否かを判定する処理部である。たとえば、判定部260は、出力値が閾値以上である場合に、音声データに異常な会話状況が含まれていると判定する。

【0029】

上記のように、本実施例1に係る学習装置100は、学習用の音声データから抽出した量子化結果と、正解情報との組を用いて、LSTMモデル110cおよびDNNモデル110dを機械学習する。このため、特定の会話状況を検出するためのキーワードを設定するための試行錯誤、熟練の知識、ノウハウを用いることなく、LSTMモデル110cおよびDNNモデル110dを機械学習することができる。また、判定装置200が、学習済みのLSTMモデル110cおよびDNNモデル110dを用いて、音声データに対する処理を行うことで、特定の会話状況が音声データに含まれているか否かを適切に判定することができる。

【0030】

[システム構成]

次に、本実施例1に係るシステムの一例について説明する。図3は、本実施例1に係るシステムの一例を示す図である。図3に示すように、このシステムは、顧客端末10と、オペレータ端末15と、通話録音装置30と、管理者端末40と、学習装置100と、判定装置200とを有する。

【0031】

顧客端末10と、オペレータ端末15とは、IP（Internet Protocol）網等のネットワーク1を介して相互に接続される。また、オペレータ端末15、通話録音装置30、管理者端末40、学習装置100、判定装置200も所定のネットワークにより、相互に接続される。なお、各ネットワークには、有線や無線を問わず、インターネットや専用線などの各種通信網を採用することができる。

【0032】

顧客端末10は、顧客がオペレータと会話（通話）するために利用する端末装置である。オペレータ端末15は、オペレータが顧客と会話するために利用する端末装置である。

【0033】

通話録音装置30は、顧客端末10と、オペレータ端末15との間で送受信される会話の音声を録音する装置である。学習時において、通話録音装置30が録音した音声データは、学習装置100に通知され、学習用の音声データとして用いられる。異常会話の検出時において、通話録音装置30が録音した音声データは、判定装置200に通知され、音声データに異常な会話状況が含まれるか否かが判定される。

【0034】

管理者端末40は、オペレータ端末15を用いて、顧客と会話するオペレータを管理する管理者が利用する端末装置である。たとえば、判定装置200が、顧客とオペレータとの会話に、異常な会話状況が含まれると判定した場合に、判定装置により、異常な会話状況を検出した旨の情報が、管理者端末40に通知される。

【0035】

10

20

30

40

50

学習装置１００は、学習用の音声データと正解情報とを用いて、ＬＳＴＭモデル１１０ｃ、ＤＮＮモデル１１０ｄを学習する装置である。学習装置１００は、学習したＬＳＴＭモデル１１０ｃ、ＤＮＮモデル１１０ｄの情報を、判定装置２００に通知する。

【００３６】

判定装置２００は、学習装置１００から通知されるＬＳＴＭモデル１１０ｃ、ＤＮＮモデル１１０ｄを用いて、顧客端末１０と、オペレータ端末１５との会話に、異常な会話状況が含まれるか否かを判定する装置である。判定装置２００は、顧客とオペレータとの会話に、異常な会話状況が含まれると判定した場合、異常な会話状況を検出した旨の情報を、管理者端末４０に通知する。

【００３７】

[学習装置の機能構成]

次に、図３に示した学習装置１００の構成の一例について説明する。図４は、本実施例１に係る学習装置の構成を示す機能ブロック図である。図４に示すように、この学習装置１００は、通信部１０１、入力部１０２、表示部１０３、記憶部１０４、制御部１０５を有する。

【００３８】

通信部１０１は、通話録音装置３０、判定装置２００とデータ通信を実行する処理部である。後述する制御部１０５は、通信部１０１を介して、通話録音装置３０、判定装置２００とデータをやり取りする。通信部１０１は、通信装置の一例である。

【００３９】

入力部１０２は、学習装置１００に各種の情報を入力するための入力装置である。入力部１０２は、キーボードやマウス、タッチパネル等に対応する。

【００４０】

表示部１０３は、制御部１０５から出力される情報を表示する装置である。表示部１０３は、液晶ディスプレイやタッチパネル等に対応する。

【００４１】

記憶部１０４は、学習用音声データベース１１０ａ、正解情報１１０ｂ、ＬＳＴＭモデル１１０ｃ、ＤＮＮモデル１１０ｄを有する。記憶部１０４は、ＲＡＭ（Random Access Memory）、フラッシュメモリ（Flash Memory）などの半導体メモリ素子や、ＨＤＤ（Hard Disk Drive）などの記憶装置に対応する。

【００４２】

学習用音声データベース１１０ａは、学習用の複数の音声データを格納するデータベースである。学習用音声データベース１１０ａに格納される各音声データは、顧客とオペレータとの会話の音声データである。

【００４３】

正解情報１１０ｂは、学習用音声データベース１１０ａに格納された各音声データに対して付与された、異常な会話状況が含まれるか否かを示す情報である。

【００４４】

ＬＳＴＭモデル１１０ｃは、第１ネットワーク（ＬＳＴＭ）のパラメータに対応する情報である。ＤＮＮモデル１１０ｄは、第２ネットワーク（ＤＮＮ）のパラメータに対応する情報である。ＬＳＴＭモデル１１０ｃ、ＤＮＮモデル１１０ｄは、学習部１６０により、機械学習される。

【００４５】

制御部１０５は、取得部１０５ａ、通知部１０５ｂ、生成部１２０、第１計算部１３０、第３計算部１４０、第２計算部１５０、学習部１６０を有する。制御部１０５は、ＣＰＵ（Central Processing Unit）やＭＰＵ（Micro Processing Unit）などによって実現できる。また、制御部１０５は、ＡＳＩＣ（Application Specific Integrated Circuit）やＦＰＧＡ（Field Programmable Gate Array）などのハードワイヤードロジックによっても実現できる。

【００４６】

10

20

30

40

50

取得部 105a は、通話録音装置 30 から、学習用音声データベース 110a の情報を取得する処理部である。取得部 105a は、学習用音声データベース 110a の情報を、記憶部 104 に格納する。また、取得部 105a は、正解情報 110b を取得した場合には、取得した正解情報 110b を、記憶部 104 に格納する。正解情報 110b は、学習用音声データベース 110a の各音声データに予め対応付けられていてもよい。

【0047】

通知部 105b は、学習済みの LSTM モデル 110c および学習済みの DNN モデル 110d を、判定装置 200 に通知する処理部である。

【0048】

生成部 120 は、学習用音声データベース 110a から学習用の音声データを取得し、音声データを基にして、量子化結果の情報を生成する処理部である。図 5 は、本実施例 1 に係る生成部を説明するための図である。図 5 に示すように、この生成部 120 は、音響処理部 121 と、量子化テーブル 122 と、ベクトル量子化部 123 と、ベクトル化部 124 とを有する。

【0049】

音響処理部 121 は、音声データから音声認識に用いる情報を抽出する処理部である。音声データから抽出される情報は、特徴量と呼ばれる。音響処理部 121 は、音声データに、32ms 程度のフレームと呼ばれる短区間を設定し、10ms 程度シフトさせながら特徴量を抽出する。たとえば、音響処理部 121 は、MFCC (Mel-Frequency Cepstrum Coefficients) を基にして、音声データから特徴量を抽出する。音響処理部 121 は、特徴量を、ベクトル量子化部 123 に出力する。

【0050】

量子化テーブル 122 は、音声データの量子化に用いるベクトルテーブルである。この量子化テーブル 122 は、発話データの特徴量の複数の代表点であって、量子化テーブル 122 の適応処理用の各音声データを用いて、量子化テーブル生成部 125 で事前に最適化される。

【0051】

(量子化テーブル 122 の生成)

ここで、量子化テーブル 122 の生成について説明する。図 6 は、量子化テーブル 122 を生成する量子化テーブル生成部 125 を説明する図である。図 6 に示すように、学習装置 100 の生成部 120 の量子化テーブル生成部 125 は、量子化テーブル 122 の最適化を行う場合に、適応処理用の各音声データに音響解析を実行して、音響処理部 121 で特徴量を生成する。そして、量子化テーブル生成部 125 は、各音声データから生成された各特徴量と量子化テーブル 122 とを突合させて量子化結果を蓄積し、量子化誤差が最小になるように量子化テーブル 122 の更新を繰り返すことにより、量子化テーブル 122 を最適化する。

【0052】

図 7 は、量子化テーブル生成部 125 で行う量子化テーブル 122 の適応制御の全体的な流れを説明する図である。ここでは、2次元の特徴量としている。図 7 に示すように、量子化テーブル生成部 125 は、白色雑音などから量子化テーブル 122 の初期値を生成し、適応処理用の音声データの量子化を行い、量子化テーブル 122 の各ベクトルを選択した特徴量の平均を用いてベクトルを更新し、この更新を繰り返すことにより、量子化誤差を削減する。つまり、量子化テーブル生成部 125 は、音声データの物理的な特徴量の分布に合わせ、量子化誤差が最小になるように、量子化テーブル 122 の更新を繰り返す。

【0053】

次に、図 8 から図 12 を用いて、量子化テーブル生成部 125 における量子化テーブル 122 の生成の詳細に説明する。図 8 は、量子化テーブル 122 の初期値を説明する図である。図 8 に示すように、量子化テーブル生成部 125 は、白色雑音などのように、周波数軸上のパワーの偏りが小さい音声データの特徴量の組から所定数 (例えば 256 個) の

量子化点を生成する。具体的には、量子化テーブル生成部125は、十分に長い白色雑音の特微量の組を算出し、その結果から順番に選択することで量子化点 $v_1(x, y)$ 、 $v_2(x, y) \cdots v_{256}(x, y)$ のそれぞれを生成し、それぞれに対して量子化番号1から量子化番号256を対応付ける。また、各量子化点は、ここでは大きさと向きを有するベクトルとする。したがって、各量子化点を、 x y 軸を有する2次元の特微量空間上で表現すると、方向や大きさが異なる256個のベクトルで表現できる。

【0054】

続いて、量子化テーブル生成部125は、適応処理用音声データから特微量を生成し、特微量を示すベクトルと、各量子化番号に対応する量子化点との距離（例えばユークリッド距離など）を算出し、距離が最も近い量子化点を選択する。ここで、量子化テーブル生成部125は、初期の量子化テーブル122で最も選択されたベクトルを無音相当とみなして他の適応に用いず、無音に相当するベクトルが1本になるように適応する。また、量子化テーブル生成部125は、各特微量の適応中に、各ベクトルを選択した特微量の平均を用いてベクトルを更新する一方で、量子化で比較的选择されなかった（例えば下位1/4など）ベクトルは更新せずに棄却し、替わりのベクトルを作成して置き換える。このようにして、量子化テーブル生成部125は、無音以外に相当する各ベクトルの選択数に大きな偏りが無いように適応させる。

【0055】

図9は、量子化点の選択を説明する図である。図9に示すように、量子化テーブル生成部125は、一連の会話データに対して、窓関数の適用や高速フーリエ変換などを実行して、所定区間に分割した各音声データから各特微量を生成する。続いて、量子化テーブル生成部125は、各特微量について、量子化テーブル122の各量子化点との距離を算出し、最も距離が近い量子化点を選択する。このようにして、量子化テーブル生成部125は、各量子化点の選択回数を計数する。

【0056】

そして、量子化テーブル生成部125は、量子化点の更新を実行する。図10は、量子化テーブル122の更新を説明する図であり、図10には、初期の量子化テーブル122における各量子化点の選択結果と、当該量子化点を選択した音声データの特微量とを図示している。

【0057】

具体的には、量子化テーブル生成部125は、選択回数が多い上位数個の量子化点と、選択回数が少ない下位数個の量子化点とを特定する。そして、量子化テーブル生成部125は、下位の量子化点を棄却（破棄）し、当該下位の量子化点の代わりに、上位の量子化点を挿入する。一方、量子化テーブル生成部125は、下位の量子化点以外の量子化点については、その量子化点を選択した音声の特微量に基づき更新する。ただし、量子化テーブル生成部125は、無音相当の量子化点を入れ替え対象から除外する。

【0058】

例えば、量子化テーブル生成部125は、選択回数が最も多い量子化番号 n の量子化点（ $v_n(x, y)$ ）を、無音に相当すると判定し、この量子化点を入れ替え対象から除外する。続いて、量子化テーブル生成部125は、選択回数が上位数個の量子化点として、量子化番号1、2などを特定するとともに、選択回数が下位数個の量子化点として、量子化番号3、256などを特定する。そして、量子化テーブル生成部125は、選択回数が少ない量子化番号3の量子化点（ $v_3(x, y)$ ）を破棄して、選択回数が多い量子化番号1の量子化点（ $v_1(x, y)$ ）を挿入し、同様に、選択回数が少ない量子化番号256の量子化点（ $v_{256}(x, y)$ ）を破棄して、選択回数が多い量子化番号2の量子化点（ $v_2(x, y)$ ）を挿入する。

【0059】

また、量子化テーブル生成部125は、選択回数が下位以外の量子化点については、その量子化点を選択した特微量の平均値で更新する。例えば、量子化テーブル生成部125は、量子化番号2の量子化点（ $v_2(x, y)$ ）について、〔選択した特微量の合計／選

10

20

30

40

50

択回数] = $[(x', y') + (x'', y'') + \dots] / 21$ により、x軸とy軸それぞれの平均値 $[v_2'(x, y)]$ を算出する。そして、量子化テーブル生成部125は、量子化番号2の「 $v_2(x, y)$ 」を「 $v_2'(x, y)$ 」に更新する。このようにして、量子化テーブル生成部125は、選択回数が少ない下位数個の量子化点以外の量子化点について、量子化番号2と同様の手法により更新する。

【0060】

図11は、量子化テーブルの更新前後を説明する図である。図11に示すように、量子化テーブル生成部125は、量子化番号1の量子化点を「 $v_1(x, y)$ 」から「 $v_1'(x, y)$ 」に更新し、量子化番号2の量子化点を「 $v_2(x, y)$ 」から「 $v_2'(x, y)$ 」に更新し、無音相当の量子化番号nの量子化点を「 $v_n(x, y)$ 」から「 $v_n'(x, y)$ 」に更新する。一方、量子化テーブル生成部125は、選択回数が少ない量子化番号3の量子化点「 $v_3(x, y)$ 」を更新前の量子化番号1の量子化点「 $v_1(x, y)$ 」に更新し、量子化番号256の量子化点「 $v_{256}(x, y)$ 」を更新前の量子化番号2の量子化点「 $v_2(x, y)$ 」に更新する。

【0061】

このようにして、量子化テーブル生成部125は、一連の適応処理用会話データの各音声データを用いて、量子化テーブル122の更新を実行すると、量子化誤差を算出する。そして、量子化テーブル生成部125は、量子化誤差が閾値以上である場合、異なるまたは同じ適応処理用会話データを用いて、同様の更新処理を実行し、量子化誤差が閾値未満である場合、「効率的なテーブル」が生成できたと判定して、更新処理を終了し、その時点の量子化テーブル122を保存する。

【0062】

図12は、量子化誤差による判定手法を説明する図である。図12の例では、量子化番号1の量子化点「 $v_1(x, y)$ 」の選択回数が「 N_1 個」、量子化番号2の量子化点「 $v_2(x, y)$ 」の選択回数が「 N_2 個」、・・・、量子化番号256の量子化点「 $v_{256}(x, y)$ 」の選択回数が「 N_{256} 個」とする。このような状態において、量子化テーブル生成部125は、量子化番号1について量子化誤差「 M_1 」、量子化番号2について量子化誤差「 M_2 」、・・・、量子化番号256について量子化誤差「 M_{256} 」を算出する。

【0063】

例えば、量子化番号2を例にして説明すると、量子化テーブル生成部125は、量子化誤差「 M_2 」として、量子化点「 $v_2(x, y)$ 」と特徴量との差分の合計値である「 $(v_2(x, y) - (x', y'))^2 + (v_2(x, y) - (x'', y''))^2 + \dots$ 」を算出する。そして、量子化テーブル生成部125は、各量子化誤差の合計「 M_{all} 」として「 $M_{all} = M_1 + M_2 + M_3 + \dots + M_{256}$ 」を算出する。ただし、無音相当の量子化番号nについては除外する。つまり、「 M_{all} 」は255個の合計値である。

【0064】

続いて、量子化テーブル生成部125は、量子化誤差「 M_{all}' 」として、各量子化誤差の合計「 M_{all} 」を、各量子化点の選択回数の合計で除算した「 $M_{all} / (N_1 + N_2 + N_3 + \dots + N_{256})$ 」を算出する。ただし、無音相当の量子化番号nについては除外する。つまり、上記算出式の分母は255個の合計値である。

【0065】

そして、量子化テーブル生成部125は、量子化誤差「 M_{all}' 」が閾値以上である場合、量子化テーブル122の更新を継続し、量子化誤差「 M_{all}' 」が閾値未満である場合、量子化テーブル122の更新を終了する。このようにして、量子化テーブル生成部125は、量子化に用いるベクトルテーブルとして、量子化テーブル122を生成する。

【0066】

図5に戻り、ベクトル量子化部123は、量子化テーブル122と特徴量を照合し、各

10

20

30

40

50

量子化点と対応している各量子化結果（例えば量子化番号に対応する）を基にして、特徴量に対応する量子化結果を出力する処理を、音響処理部 121 から特徴量を受け付ける度に、実行する処理部である。ベクトル量子化部 123 は、各特徴量に対応する量子化結果を時系列に並べた量子化系列の情報を、ベクトル化部 124 に出力する。

【0067】

ベクトル化部 124 は、量子化系列に含まれる各量子化結果をベクトルに変換する処理部である。ベクトル化部は、量子化系列の各量子化結果に対応する各ベクトルを、第 1 計算部 130 に出力する。たとえば、ベクトル化部 124 は、各量子化結果を、40 次元の One Hot ベクトルで表す。入力された量子化結果の One Hot ベクトルは、入力された量子化結果の次元に「1」が設定され、他の次元に「0」が設定される。

10

【0068】

図 13 は、本実施例 1 に係るベクトル化部の処理を説明するための図である。たとえば、ベクトル化部 124 は、量子化結果 1、量子化結果 2、・・・、量子化結果 256 が順に入力された場合には、One Hot ベクトル V_1 、 V_2 、・・・、 V_m を生成する。One Hot ベクトル V_1 は、量子化結果 2 に対応する次元に「1」が設定され、他の次元には「0」が設定される。One Hot ベクトル V_2 は、量子化結果 256 に対応する次元に「1」が設定され、他の次元には「0」が設定される。One Hot ベクトル V_m は、量子化結果 4 に対応する次元に「1」が設定され、他の次元には「0」が設定される。

【0069】

図 4 の説明に戻る。第 1 計算部 130 は、再帰パスをもつ第 1 ネットワークに、各量子化結果の One Hot ベクトルを順に入力し、第 1 ネットワークのパラメータに基づく計算を行うことで、内部ベクトルを算出する処理部である。

20

【0070】

第 1 計算部 130 は、第 1 ネットワークとして LSTM を用いる。図 14 は、本実施例 1 に係る LSTM の一例を示す図である。LSTM 135 は、1 個につき、重み行列とバイアスとの組を 3 組もつ（1 層のフォワード型ニューラルネットワークが 3 つ接続されたものに対応する）。かかる重み行列の値と、バイアスの値が、LSTM 135 の学習対象のパラメータとなる。

【0071】

LSTM 135 は、 $\sigma 1$ 、 $\sigma 2$ 、 $\sigma 3$ と、スイッチ 2a、2b、2c と、加算部 2d と、 $\tanh 1$ と、 $\tanh 2$ とを有する。 $\sigma 1 \sim \sigma 3$ は、入力されたベクトルと、重み行列とを乗算することで、アフィン変換を行い、アフィン変換したベクトルを出力する処理部である。スイッチ 2a $\sim$ 2c は、入力されたベクトルの各次元の値に応じて、ゲートを通過するベクトルの次元毎に ON/OFF を制御する処理部である。たとえば、スイッチ 2a $\sim$ 2c は、シグモイド関数等を基にして、ON/OFF を制御する。加算部 2d は、2 方向から入力されたベクトルを次元毎に加算した値を出力する処理部である。 $\tanh 1$ 、 $\tanh 2$ は、入力されたベクトルに対して、 $\tanh$ 関数に基づく計算を行い、計算結果を出力する処理部である。

30

【0072】

図 14 において、「 X_t 」は、時刻 t の量子化結果の One Hot ベクトルを示す。便宜上、時刻 $t-1$ において、量子化結果の One Hot ベクトル「 X_{t-1} 」を入力した時点の LSTM 135 を LSTM 135 $_{t-1}$ と表記する。時刻 t において、量子化結果の One Hot ベクトル「 X_t 」を入力した時点の LSTM 135 を LSTM 135 $_t$ と表記する。時刻 $t+1$ において、量子化結果の One Hot ベクトル「 X_{t+1} 」を入力した時点の LSTM 135 を LSTM 135 $_{t+1}$ と表記する。

40

【0073】

一例として、LSTM 135 $_t$ を用いて説明を行う。「 X_t 」が入力されると、LSTM 135 $_{t-1}$ から入力された h_{t-1} と、 X_t とを加算したベクトルが、 $\sigma 1$ 、 $\sigma 2$ 、 $\sigma 3$ と、 $\tanh 1$ に入力される。 h_{t-1} は、時刻 $t-1$ において、LSTM 135 に算出される内部ベクトルである。

50

【0074】

スイッチ2aは、 $\sigma 1$ から出力されるベクトルを基にして、 S_{t+1} が通過するゲートのON/OFFを制御する。スイッチ2bは、 $\sigma 2$ から出力されるベクトルを基にして、 $\tanh 1$ から出力されるベクトルが通過するゲートのON/OFFを制御する。スイッチ2cは、 $\sigma 3$ から出力されるベクトルを基にして、加算部2dから出力されるベクトル S_t が通過するゲートのON/OFFを制御する。スイッチ2cから出力されるベクトルが、時刻tにおける内部ベクトル h_t となる。内部ベクトル h_t は、LSTM135 $_{t+1}$ に入力される。

【0075】

加算部2dは、スイッチ2aから出力されるベクトルと、スイッチ2bから出力されるベクトルとを加算したベクトル S_t を算出する処理部である。ベクトル S_t は、LSTM135 $_t$ に入力されると共に、 $\tanh 2$ に入力される。

10

【0076】

第1計算部130は、図14に説明したLSTM135に、量子化系列に含まれる全量子化結果のOne Hotベクトルをそれぞれ順に入力することで、複数の内部ベクトルhを算出する。第1計算部130は、複数の内部ベクトルhを、第3計算部140に出力する。

【0077】

第3計算部140は、第1計算部130から出力される複数の内部ベクトルhを平均化する処理部である。第3計算部140は、平均化した内部ベクトル（平均ベクトル）を、第2計算部150に出力する。

20

【0078】

第2計算部150は、再帰パスを持たない第2ネットワークに、平均ベクトルを入力し、第2ネットワークのパラメータに基づく計算を行うことで、出力値（ニューロン値）を算出する処理部である。

【0079】

たとえば、第2ネットワークは、フィードフォワード型のニューラルネットワークである。図15は、本実施例1に係るニューラルネットワークの一例を説明するための図である。図15に示すように、このニューラルネットワーク155は、入力層20a、隠れ層20b、出力層20cを持つ。入力層20a、隠れ層20b、出力層20cは、複数のノードがエッジで結ばれる構造となっている。隠れ層20b、出力層20cは、活性化関数と呼ばれる関数とバイアス値とを持ち、エッジは、重みを持つ。かかるバイアス値、重みが、第2ネットワークの学習対象となるパラメータとなる。

30

【0080】

入力層20aに含まれる各ノードに、平均ベクトルを入力すると、隠れ層20bを通して、出力層20cの各ノードから、会話が異常な会話状況である確率「Ot」と、会話が通常の会話状況である確率「On」とが出力される。たとえば、出力層20cは、softmaxにより確率化され、「Ot」の出力値と「On」の出力値との和が「1.0」になる。

【0081】

図4の説明に戻る。学習部160は、音声データから生成された各量子化結果のOne Hotベクトルを第1計算部130に入力した際に、第3計算部140を介して、第2計算部150から出力される出力値が、音声データに対応する正解情報110bに近づくように、第1計算部130のパラメータ、第2計算部150のパラメータを学習する処理部である。

40

【0082】

たとえば、学習部160は、正解情報「異常な会話状態」に対応する音声データから生成される各量子化結果のOne Hotベクトルを第1計算部130に入力する場合、確率「Ot」が「1」に近づき、確率「On」が「0」に近づくように、パラメータを学習する。学習部160は、正解情報「正常な会話状態」に対応する音声データから生成される各量子化結果のOne Hotベクトルを第1計算部130に入力する場合、確率「Ot」が「0」に近づき、確率「On」が「1」に近づくように、パラメータを学習する。

50

【0083】

学習部160は、O t から出力される値と正解の値との差分、および、O n から出力される値と正解の値との差分を含む損失関数として、たとえば、Cross Entropyを用いる。学習部160は、損失関数の誤差を逆伝播することで、損失関数の値が最小値となるように、パラメータの学習を繰り返し実行する。学習部160は、学習停止条件を設定し、学習停止条件を満たす場合に、学習を終了する。たとえば、学習停止条件は、損失関数の値が閾値未満となる等の条件である。

【0084】

学習部160は、LSTM135のパラメータの学習結果の情報を、LSTMモデル110cとして、記憶部104に格納する。学習部160は、ニューラルネットワーク155のパラメータの学習結果の情報を、DNNモデル110dとして、記憶部104に格納する。

10

【0085】

[判定装置の機能構成]

次に、図3に示した判定装置200の構成の一例について説明する。図16は、本実施例1に係る判定装置の構成を示す機能ブロック図である。図16に示すように、判定装置200は、通信部201、入力部202、表示部203、記憶部204、制御部205を有する。図示を省略するが、判定装置200は、音声データを取得するためのマイクに接続されていてもよい。

【0086】

通信部201は、通話録音装置30、学習装置100とデータ通信を実行する処理部である。後述する制御部205は、通信部201を介して、通話録音装置30、学習装置100とデータをやり取りする。通信部201は、通信装置の一例である。

20

【0087】

入力部202は、判定装置200に各種の情報を入力するための入力装置である。入力部202は、キーボードやマウス、タッチパネル等に対応する。

【0088】

表示部203は、制御部205から出力される情報を表示する装置である。表示部203は、液晶ディスプレイやタッチパネル等に対応する。

【0089】

記憶部204は、音声データ204a、LSTMモデル110c、DNNモデル110dを有する。記憶部204は、RAM、フラッシュメモリなどの半導体メモリ素子や、HDDなどの記憶装置に対応する。

30

【0090】

音声データ204aは、通話録音装置30によって録音された顧客とオペレータとの会話の音声データであって、異常な会話状況の検出対象となる音声データである。

【0091】

LSTMモデル110cは、学習装置100によって学習された、第1ネットワーク（LSTM135）のパラメータに対応する情報である。

【0092】

DNNモデル110dは、学習装置100によって学習された、第2ネットワーク（ニューラルネットワーク155）のパラメータに対応する情報である。

40

【0093】

制御部205は、取得部205a、通知部205b、生成部220、第1計算部230、第3計算部240、第2計算部250、判定部260を有する。制御部205は、CPUやMPUなどによって実現できる。また、制御部205は、ASICやFPGAなどのハードワイヤードロジックによっても実現できる。

【0094】

取得部205aは、通話録音装置30から、音声データ204aを取得する処理部である。取得部205aは、音声データ204aを記憶部204に格納する。また、取得部2

50

05aは、学習装置100から、学習済みのLSTMモデル110cの情報、および、DNNモデル110dの情報を取得する。取得部205aは、LSTMモデル110cの情報、および、DNNモデル110dの情報を、記憶部204に格納する。

【0095】

通知部205bは、後述する判定部260から判定結果を取得する。通知部205bは、判定部260によって、音声データ204aに異常な会話状況が含まれていると判定された場合に、異常な会話が含まれる旨の情報を、管理者端末40に通知する。

【0096】

生成部220は、音声データ204aを取得し、音声データ204aを基にして、量子化系列の情報を生成する処理部である。生成部220は、学習装置100の生成部120と同様にして、音声データ204aから量子化系列を生成し、各量子化結果のOne Hotベクトルを生成する。生成部220は、各量子化結果のOne Hotベクトルを、第1計算部230に出力する。

10

【0097】

第1計算部230は、再帰パスをもつ第1ネットワークに、各量子化結果の各量子化結果のOne Hotベクトルを順に入力し、第1ネットワークのパラメータに基づく計算を行うことで、内部ベクトルを算出する処理部である。第1計算部230が用いる第1ネットワークは、図14で説明したLSTM135に対応するものである。第1計算部230は、LSTM135のパラメータに、LSTMモデル110cのパラメータを設定し、内部ベクトルを計算する。第1計算部230は、各量子化結果のOne Hotベクトルから算出される内部ベクトルhを、第2計算部250に出力する。

20

【0098】

第3計算部240は、第1計算部230から出力される複数の内部ベクトルhを平均化する処理部である。第3計算部240は、平均化した内部ベクトル（平均ベクトル）を、第2計算部250に出力する。

【0099】

第2計算部250は、再帰パスを持たない第2ネットワークに、平均ベクトルを入力し、第2ネットワークのパラメータに基づく計算を行うことで、出力値（ニューロン値）を算出する処理部である。第2計算部250が用いる第2ネットワークは、図15で説明したニューラルネットワーク155に対応するものである。第2計算部250は、ニューラルネットワーク155のパラメータに、DNNモデル110dのパラメータを設定し、会話が異常な会話状況である確率「Ot」と、会話が通常の会話状況である確率「On」とを計算する。第2計算部250は、確率「Ot」の情報を、判定部260に出力する。

30

【0100】

判定部260は、異常な会話状況である確率「Ot」を基にして、音声データ204aに異常な会話状況が含まれているか否かを判定する処理部である。たとえば、判定部260は、確率Otが、予め設定される閾値以上となる場合に、音声データ204aに異常な会話状況が含まれていると判定する。判定部260は、判定結果を、通知部205bに出力する。この例では、異常な会話状況である確率「Ot」を基にして判定しているが、「Ot」と正常な会話状況である確率「On」の確率との差または比率を判定に用いる値として用いてもよい。

40

【0101】

〔量子化テーブルの生成処理の流れ〕

図17は、本実施例1にかかる量子化テーブル122の生成処理の流れを示すフローチャートである。図17に示すように、生成部120は、処理開始が指示されると（S1：Yes）、量子化テーブル122の初期化を実行する（S2）。

【0102】

続いて、生成部120は、適応処理用会話データを取得し（S3）、音声データに分割する（S4）。そして、生成部120は、1つの音声データを取得し（S5）、音声データの特徴量を算出する（S6）。

50

【0103】

その後、生成部120は、ベクトル量子化により選択を実行し（S7）、選択結果を蓄積する（S8）。すなわち、生成部120は、量子化テーブル122の各量子化点と、音声データの特徴量との距離を算出し、当該距離が一番短い量子化点を選択する。

【0104】

ここで、生成部120は、未処理の音声データが存在する場合（S9：Yes）、次の音声データについてS5以降を繰り返す。一方、生成部120は、未処理の音声データが存在しない場合（S9：No）、量子化テーブル122の選択結果に基づき、無音に相当する量子化点（量子化結果）を特定する（S10）。なお、無音がすでに特定済みの場合、S10は省略される。

10

【0105】

続いて、生成部120は、選択回数が多い量子化点と選択回数が少ない量子化点を特定し（S11）、量子化テーブル122を更新する（S12）。すなわち、生成部120は、選択回数が少ない量子化点を、無音に相当する量子化点以外で選択回数が多い更新前の量子化点に変更し、その他の量子化点を特徴量の平均に更新する。

【0106】

その後、生成部120は、量子化誤差を算出し（S13）、量子化誤差が閾値以上の場合（S14：Yes）、異なるまたは同じ会話データに対してS3以降を繰り返す。一方、生成部120は、量子化誤差が閾値未満の場合（S14：No）、処理を終了し、量子化テーブル122を出力する（S15）。

20

【0107】

[学習処理の流れ]

次に、本実施例1に係る学習装置100の処理手順の一例について説明する。図18は、本実施例1に係る学習装置の処理手順を示すフローチャートである。図18に示すように、学習装置100の生成部120は、学習用の音声データを取得して、ベクトル量子化を行う（ステップS101）。生成部120は、量子化結果をOne-hotベクトル化する（ステップS102）。生成部120は、One-hotベクトルを蓄積する（ステップS103）。

【0108】

生成部120は、学習データ数（One-hotベクトルの数）が、モデル学習可能な基準を超えていない場合には（ステップS104、No）、ステップS101に移行する。一方、生成部120は、学習データ数が、モデル学習可能な基準を超えた場合には（ステップS104、Yes）、ステップS105に移行する。

30

【0109】

学習装置100の第1計算部130は、One-hotベクトルをLSTM135に入力し、内部ベクトルを算出し、蓄積する（ステップS105）。第1計算部130は、全One-hotベクトル系列を入力していない場合には（ステップS106、No）、ステップS105に移行する。一方、第1計算部130は、全One-hotベクトル系列を入力した場合には（ステップS106、Yes）、ステップS107に移行する。学習装置100の第3計算部140は、内部ベクトルを平均化する（ステップS107）。

40

【0110】

学習装置100の第2計算部150は、平均化した内部ベクトルをニューラルネットワーク155に入力し、出力値を算出する（ステップS108）。学習装置100の学習部160は、正解情報110bを用いてLSTM135およびニューラルネットワーク155のパラメータを誤差逆伝播法によって学習する（ステップS109）。

【0111】

学習部160は、全学習データについて学習を行っていない場合には（ステップS110、No）、ステップS105に移行する。学習部160は、全学習データについて学習を行った場合には（ステップS110、Yes）、ステップS111に移行する。

【0112】

50

学習部 160 は、学習停止条件を満たしていない場合には（ステップ S111, No）、学習データの順番をランダムに入れ替えた後に、ステップ S105 に移行する。学習部 160 は、学習停止条件を満たした場合には（ステップ S111, Yes）、ステップ S112 に移行する。学習部 160 は、学習済みの LSTM モデル 110c および DNN モデル 110d を、記憶部 104 に格納する（ステップ S112）。

【0113】

[判定処理の流れ]

次に、本実施例 1 に係る判定装置 200 の処理手順の一例について説明する。図 19 は、本実施例 1 に係る判定装置の処理手順を示すフローチャートである。図 19 に示すように、判定装置 200 の生成部 220 は、ベクトル量子化を行い（ステップ S201）、量子化結果を One-hot ベクトル化する（ステップ S202）。生成部 220 は、One-hot ベクトルを蓄積する（ステップ S203）。生成部 220 は、全量子化結果を One-hot ベクトル化していない場合には（ステップ S204, No）、ステップ S202 に移行する。一方、生成部 220 は、全量子化結果を One-hot ベクトル化した場合には（ステップ S204, Yes）、ステップ S205 に移行する。

【0114】

判定装置 200 の第 1 計算部 230 は、One-hot ベクトルを LSTM 135 に入力し、内部ベクトルを算出し、蓄積する（ステップ S205）。なお、第 1 計算部 230 は、学習済みの LSTM モデル 110c に基づく LSTM 135 を用いて内部ベクトルを算出する。第 1 計算部 230 は、全 One-hot ベクトル系列を入力していない場合には（ステップ S206, No）、ステップ S205 に移行する。第 1 計算部 230 は、全 One-hot ベクトル系列を入力した場合には（ステップ S206, Yes）、ステップ S207 に移行する。

【0115】

判定装置 200 の第 3 計算部 240 は、内部ベクトルを平均化する（ステップ S207）。判定装置 200 の第 2 計算部 250 は、平均化した内部ベクトルをニューラルネットワーク 155 に入力し、出力値「Ot」を算出する（ステップ S208）。なお、第 2 計算部 250 は、学習済みの DNN モデル 110d に基づくニューラルネットワーク 155 を用いて、出力値を算出する。「Ot」は、会話が異常な会話状況である確率を示すものである。

【0116】

判定装置 200 の判定部 260 は、出力値 Ot が閾値以上であるか否かを判定する（ステップ S209）。判定部 260 は、出力値 Ot が閾値以上である場合には（ステップ S209, Yes）、音声データ 204a に異常な会話状況が含まれていると判定する（ステップ S210）。判定装置 200 の通知部 205b は、異常な会話状況が含まれている旨を管理者端末 40 に通知する（ステップ S211）。

【0117】

一方、判定部 260 は、出力値 Ot が閾値未満である場合には（ステップ S209, No）、通常会話であると判定する（ステップ S212）。

【0118】

[効果]

次に、本実施例 1 に係る学習装置 100 の効果について説明する。学習装置 100 は、学習用の音声データから抽出した量子化系列と、正解情報との組を用いて、LSTM モデル 110c および DNN モデル 110d を機械学習する。このため、特定の会話状況を検出するためのキーワードを設定するための試行錯誤、熟練の知識、ノウハウを用いることなく、LSTM モデル 110c および DNN モデル 110d を機械学習することができる。また、判定装置 200 が、学習済みの LSTM モデル 110c および DNN モデル 110d を用いて、音声データに対する処理を行うことで、特定の会話状況が音声データに含まれているか否かを適切に判定することができる。

【0119】

10

20

30

40

50

図20は、異常な会話状況が含まれる音声データの内部ベクトルの一例を示す図である。図20のグラフ3Aの縦軸は内部ベクトルの値を示し、横軸は内部ベクトルの次元を示す。線分3aは、異常な会話状況が含まれる第1通話の音声データの内部ベクトルを示す。線分3bは、異常な会話状況が含まれる第2通話の音声データの内部ベクトルを示す。異常な会話状況が含まれる。図20に示すように、異常状態の通話では、別通話であっても、内部ベクトルの形状は同じような形状となる。

【0120】

図21は、正常な会話の音声データの内部ベクトルの一例を示す図である。図21のグラフ3Bの縦軸は内部ベクトルの値を示し、横軸は内部ベクトルの次元を示す。線分3cは、正常な第3通話の音声データの内部ベクトルを示す。線分3dは、正常な第4通話の音声データの内部ベクトルを示す。正常な通話では、別通話であっても、内部ベクトルの形状は、同じような形状となる。

10

【0121】

図20に示した異常な会話状況の音声データのベクトル形状と、図21に示した通常の通話の音声データのベクトル形状とを比較すると、各ベクトル形状は大きく異なる。このため、内部ベクトルを用いることで、キーワードを設定することなく、異常な会話状況を含むか否かを判定でき、このためのLSTMモデル110c、DNNモデル110dを適切に学習することができる。

【0122】

また、学習装置100は、人間の聴覚特性を参考に、音声の物理的な特徴量（パワースペクトルなど）を量子化し、その結果を基にして、会話状況の分析に用いる発声の偏りを検出する。また、学習装置100は、自動調整により、無音相当を除く量子化点毎の選択数に大きな偏りが無い「効率的な量子化テーブル」を生成することができる。

20

【0123】

例えば、言語依存情報の音素のモデルなどを用いた音響モデルも考えられるが、当該音響モデルは、言語ごとに生成する必要があるため、膨大な時間とコストがかかる。一方、学習装置100は、「効率的な量子化テーブル」を用いた発声の偏りの検出により、高コストの言語依存情報のモデル生成を行うことなく、会話状況の分析を実行することができる。

【実施例2】

【0124】

さて、これまで本発明の実施例について説明したが、本発明は上述した実施例以外にも、種々の異なる形態にて実施されてよいものである。

30

【0125】

[拡張例]

第1計算部および第2計算部の拡張例について説明する。図22は、第1計算部および第2計算部の拡張例を説明するための図である。ここでは一例として、第1計算部130および第150を用いて説明する。

【0126】

第1計算部130は、第1ネットワークであるLSTMを多段にして、上段へのLSTMの内部ベクトルdを異常／通常の2次元のベクトルにする。なお、内部ベクトルcは、図14で説明した内部ベクトルhに対応するものである。たとえば、内部ベクトルcは、128次元のベクトルである。内部ベクトルdは、第3計算部140（図示略）を通過して、第2計算部150に入力される。

40

【0127】

第2計算部150は、第2ネットワークを実行しないで、内部ベクトルdに対するSoftmax計算のみを実行し、異常な会話状況である確率と、通常の会話である確率を算出する。第2計算部150の必須構成要素は、Softmax計算を行う処理部であり、内部ベクトルdから異常／通常それぞれの確からしさである確率を求めることであるため、第1計算部130、第2計算部150を図22のように拡張することが可能である。

【0128】

50

第3計算部のその他の処理について説明する。図23は、第3計算部のその他の処理を説明するための図である。ここでは一例として、第1計算部130、第2計算部150、第3計算部140を用いて説明する。

【0129】

第3計算部140は、入力量子化結果ごとに第1計算部130から出力される内部ベクトル $c_1 \sim c_N$ に対し、先頭から「 a_1 、 a_2 、 $\dots$ 、 a_N 」という重みパラメータを設ける。かかる重みは、時間方向のパラメータである。たとえば、第3計算部140から、第2計算部150に出力されるベクトルは「 $a_1 \cdot c_1 + a_2 \cdot c_2 + \dots + a_N \cdot c_N$ 」となる。学習部160は、LSTMモデル110cおよびDNNモデル110dを学習する際に、第3計算部140の重みパラメータも合わせて学習を行う。このような処理を行うことで、平均ベクトルを生成する場合の時間方向についての重みを最適化することができる。

10

【0130】

[適用例]

図24は、本実施例にかかるその他のシステムの一例を示す図である。図24では一例として、学習装置100と、判定装置200とをもちいて説明を行う。

【0131】

学習装置100は、窓口対応時の会話の学習データ180を基にして、LSTMモデル180A、DNNモデル180Bを学習する。学習データ180には、過去に行われた窓口対応時の会話を録音した会話音声録音ファイルと、トラブルが発生したか否かの正解情報を有する。窓口対応の会話は、マイク45を用いて録音される。

20

【0132】

判定装置200は、窓口の音声を取得し、LSTMモデル180A、DNNモデル180Bを用いて、トラブル発生の確からしさの情報290を算出する。判定装置200は、トラブル発生の確からしさの情報290を、管理者端末40に通知する。管理者担当者41は、トラブル発生の確からしさの情報290を参照し、対応支援を行う。

【0133】

図24に示したシステムを構築することで、窓口対応におけるトラブル発声を検出して、対応支援を行うことが可能となる。

【0134】

[数値等]

上記実施例で用いた数値例、閾値等は、あくまで一例であり、任意に変更することができる。また、本実施例では、特定の会話状況を「異常な会話状況」として説明したが、特定の会話状況は、異常な会話状況に限定されるものではない。たとえば、特定の会話状況は、会議が停滞している会話状況、トラブルが発生している会話状況、顧客にとって好ましい会話状況などであってもよい。また、量子化点についても2次元のベクトルに限らず、多次元のベクトルを用いることができる。

30

【0135】

[初期化]

上記実施例では、白色雑音を用いて量子化テーブル122の初期値を生成する例を説明したが、これに限定されるものではない。例えば、いわゆるピンクノイズや、予め用意した疑似音声などを用いることもできる。また、白色雑音やピンクノイズの特徴量からではなく、これらの雑音を用いて量子化テーブルの初期値を直接生成することもできる。

40

【0136】

[システム]

上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。なお、リアルタイム処理部33は、生成部に対応し、カウント部32は、格納制御部に対応し、要求処理部34は、取得部と生成部に対応する。

【0137】

50

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散や統合の具体的形態は図示のものに限られない。つまり、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。例えば、学習装置１００の制御部１０５は、判定装置２００の制御部２０５と同様の機能を持ち、ＬＳＴＭモデル１１０ｃ、ＤＮＮモデル１１０ｄを学習すると共に、音声データに特定の会話状況が含まれているか否かを判定してもよい。

【０１３８】

さらに、各装置にて行なわれる各処理機能は、その全部または任意の一部が、ＣＰＵおよび当該ＣＰＵにて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。

【０１３９】

[ハードウェア]

次に、学習装置１００と判定装置２００のハードウェア構成例を説明する。なお、学習装置１００と判定装置２００は、同様のハードウェア構成を有するので、ここでは、情報処理装置５００として説明する。図２５は、ハードウェア構成例を説明する図である。図２５に示すように、情報処理装置５００は、通信装置５００ａ、ＨＤＤ（Hard Disk Drive）５００ｂ、メモリ５００ｃ、プロセッサ５００ｄを有する。また、図２５に示した各部は、バス等で相互に接続される。

【０１４０】

通信装置５００ａは、ネットワークインタフェースカードなどであり、他のサーバとの通信を行う。ＨＤＤ５００ｂは、図４や図１６に示した機能を動作させるプログラムやＤＢを記憶する。

【０１４１】

プロセッサ５００ｄは、図４や図１６に示した各処理部と同様の処理を実行するプログラムをＨＤＤ５００ｂ等から読み出してメモリ５００ｃに展開することで、図４や図１６等で説明した各機能を実行するプロセスを動作させる。例えば、学習装置１００を例にして説明すると、このプロセスは、学習装置１００が有する各処理部と同様の機能を実行する。具体的には、プロセッサ５００ｄは、取得部１０５ａ、通知部１０５ｂ、生成部１２０、第１計算部１３０、第３計算部１４０、第２計算部１５０、学習部１６０等と同様の機能を有するプログラムをＨＤＤ５００ｂ等から読み出す。そして、プロセッサ５００ｄは、取得部１０５ａ、通知部１０５ｂ、生成部１２０、第１計算部１３０、第３計算部１４０、第２計算部１５０、学習部１６０等と同様の処理を実行するプロセスを実行する。

【０１４２】

このように、情報処理装置５００は、プログラムを読み出して実行することで学習方法を実行する情報処理装置として動作する。また、情報処理装置５００は、媒体読取装置によって記録媒体から上記プログラムを読み出し、読み出された上記プログラムを実行することで上記した実施例と同様の機能を実現することもできる。なお、この他の実施例でいうプログラムは、情報処理装置５００によって実行されることに限定されるものではない。例えば、他のコンピュータまたはサーバがプログラムを実行する場合や、これらが協働してプログラムを実行するような場合にも、本発明を同様に適用することができる。

【符号の説明】

【０１４３】

- １０ 顧客端末
- １５ オペレータ端末
- ３０ 通話録音装置
- ４０ 管理者端末
- １００ 学習装置
- ２００ 判定装置
- １０１，２０１ 通信部

10

20

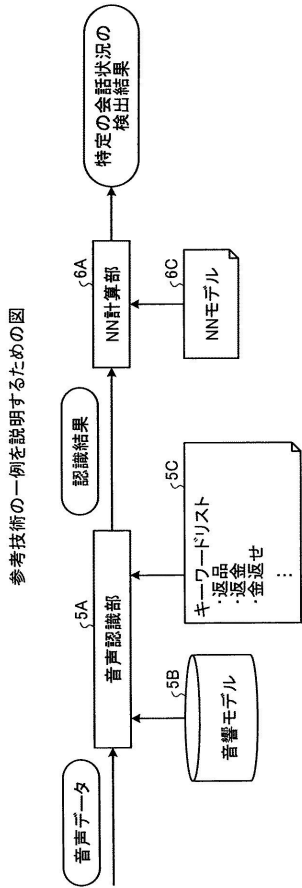
30

40

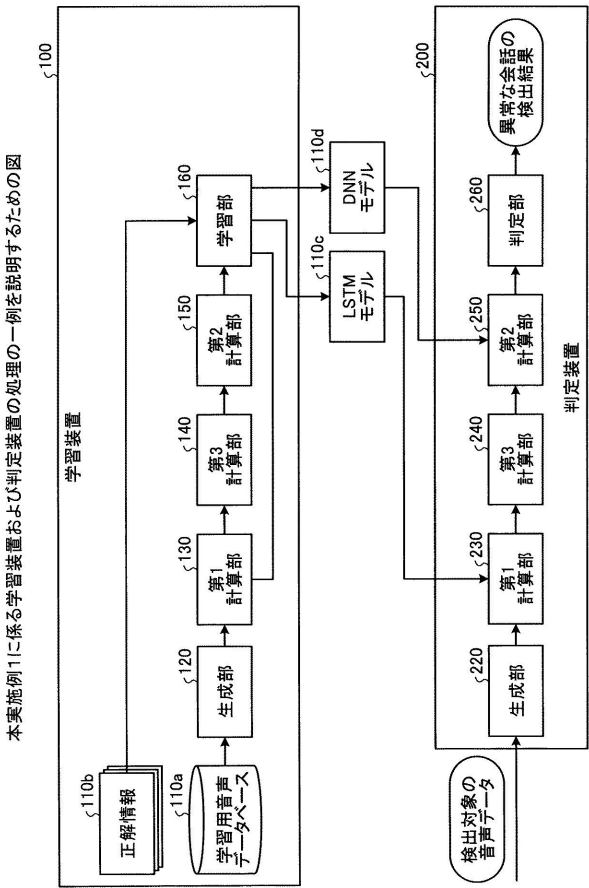
50

- 102, 202 入力部
- 103, 203 表示部
- 104, 204 記憶部
- 105, 205 制御部
- 105a, 205a 取得部
- 105b, 205b 通知部
- 110a 学習用音声データベース
- 110b 正解情報
- 110c LSTMモデル
- 110d DNNモデル
- 120, 220 生成部
- 130, 230 第1計算部
- 140, 240 第3計算部
- 150, 250 第2計算部
- 160 学習部
- 260 判定部

【図1】

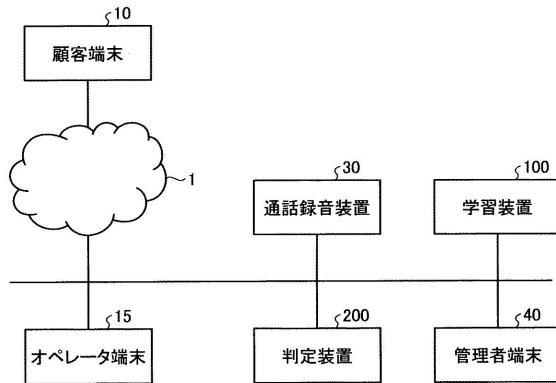


【図2】



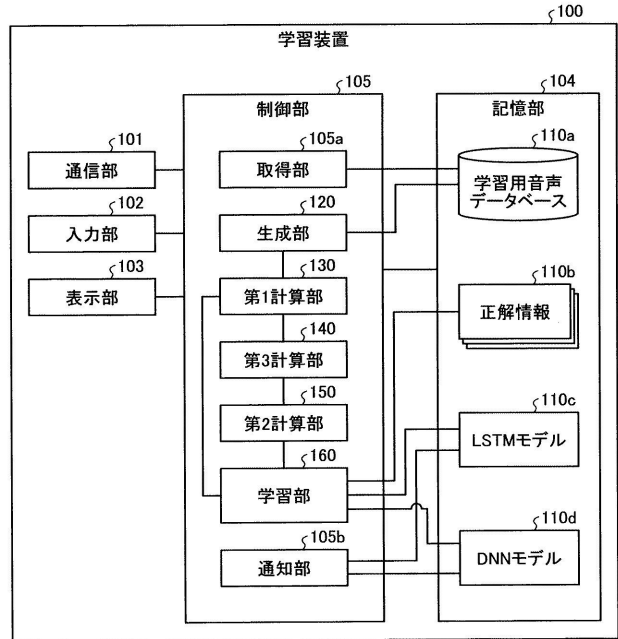
【図 3】

本実施例1に係るシステムの一例を示す図



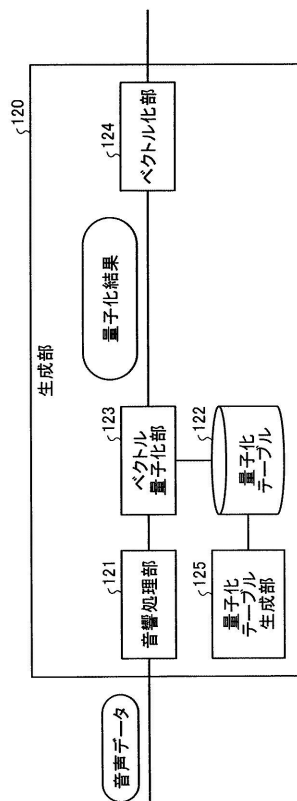
【図 4】

本実施例1に係る学習装置の構成を示す機能ブロック図



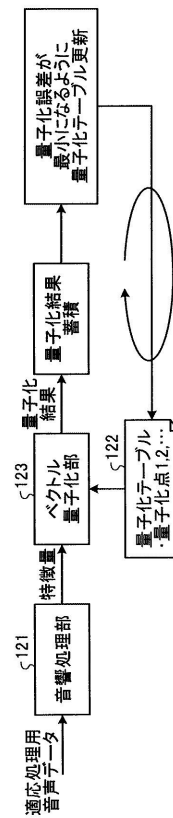
【図 5】

本実施例1に係る生成部を説明するための図

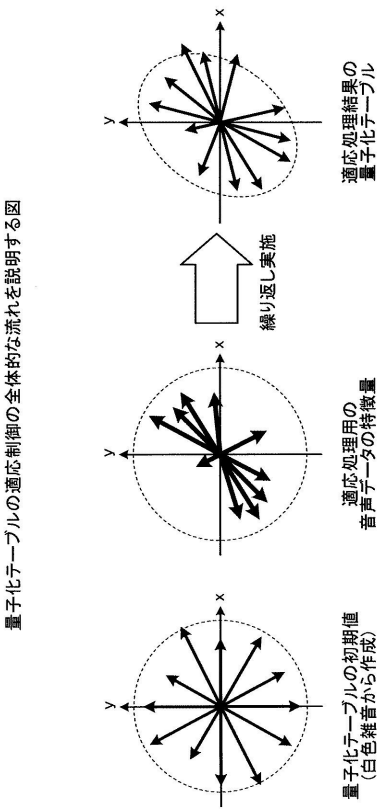


【図 6】

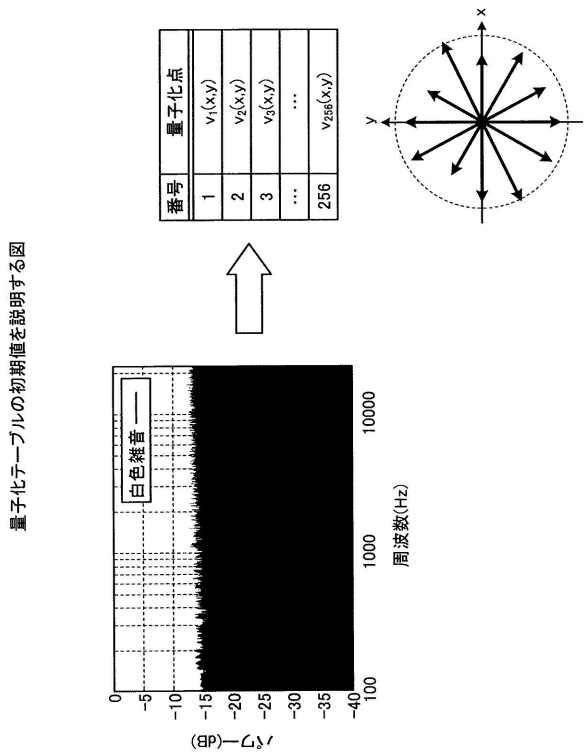
量子化テーブルを生成する量子化テーブル生成部を説明する図



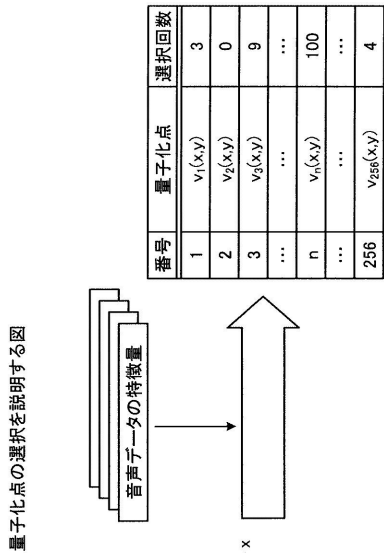
【図 7】



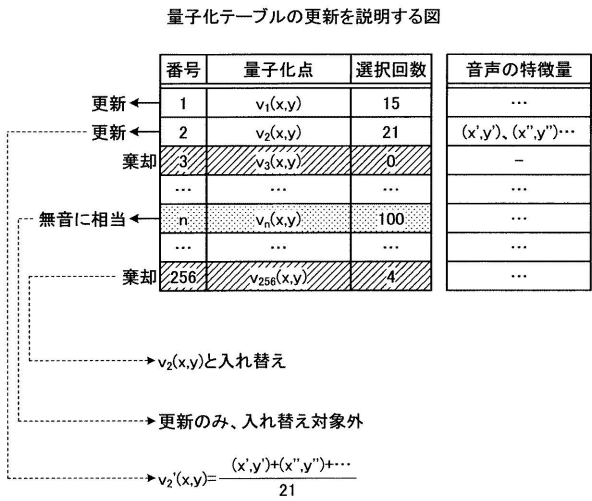
【図 8】



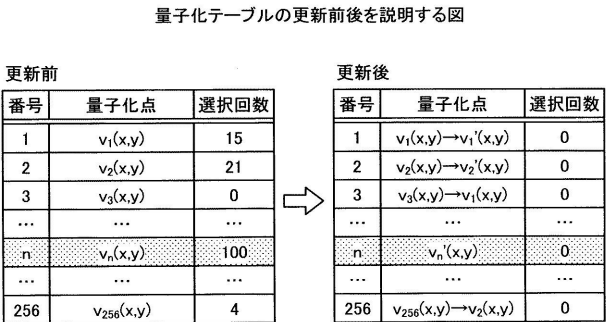
【図 9】



【図 10】

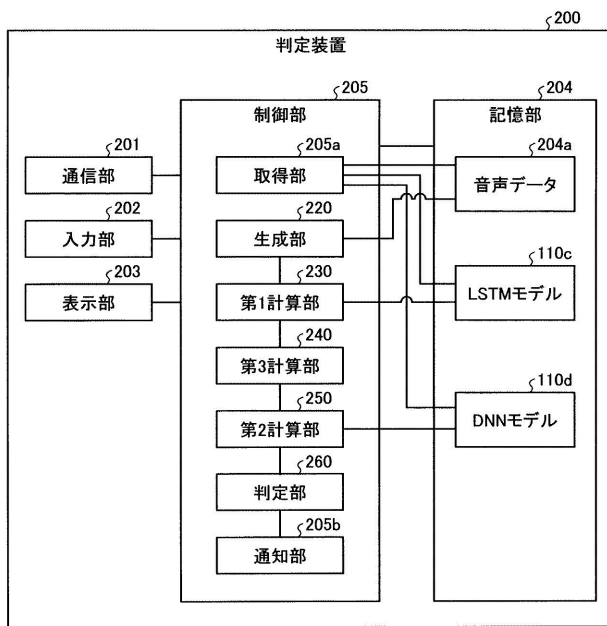


【図 11】



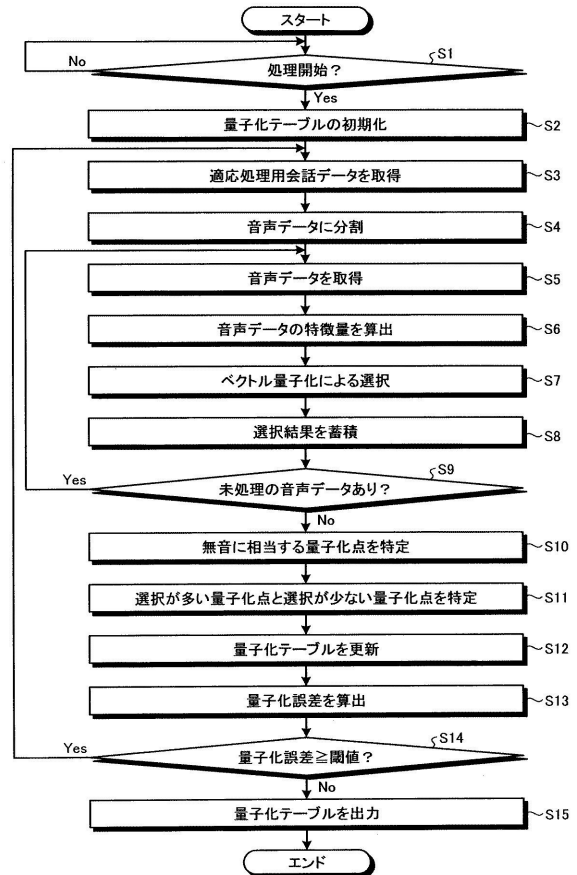
【図 16】

本実施例1に係る判定装置の構成を示す機能ブロック図



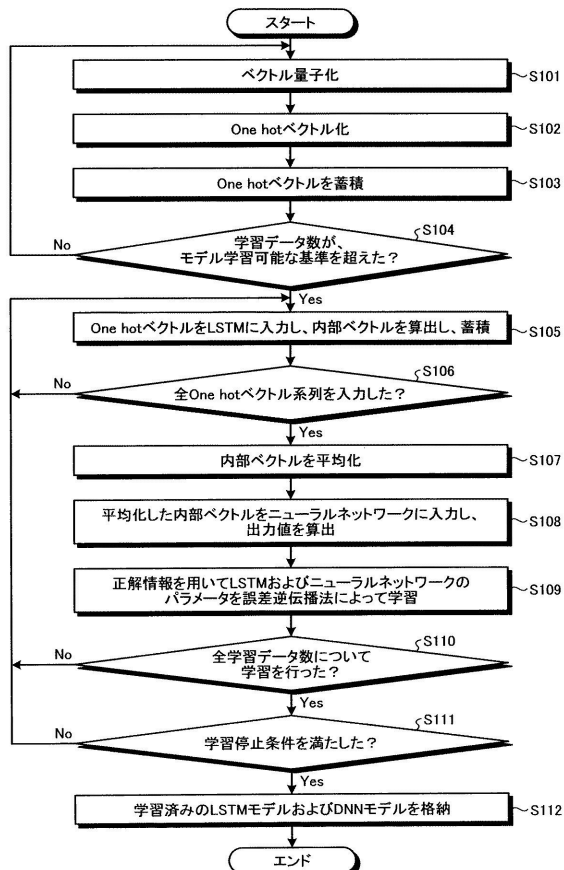
【図 17】

本実施例1に係る量子化テーブルの生成処理の流れを示すフローチャート



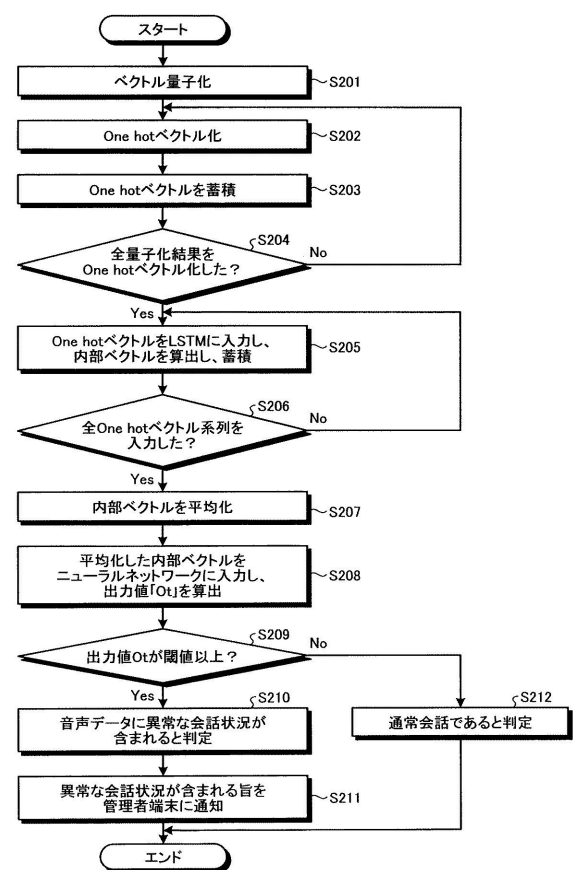
【図 18】

本実施例1に係る学習装置の処理手順を示すフローチャート



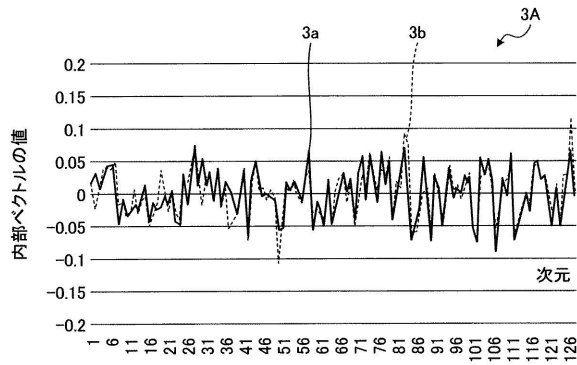
【図 19】

本実施例1に係る判定装置の処理手順を示すフローチャート



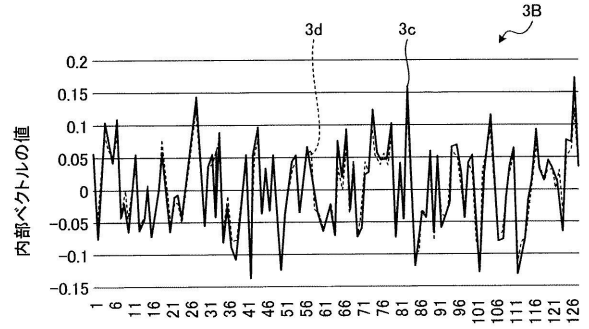
【図 20】

異常な会話状況が含まれる音声データの内部ベクトルの一例を示す図



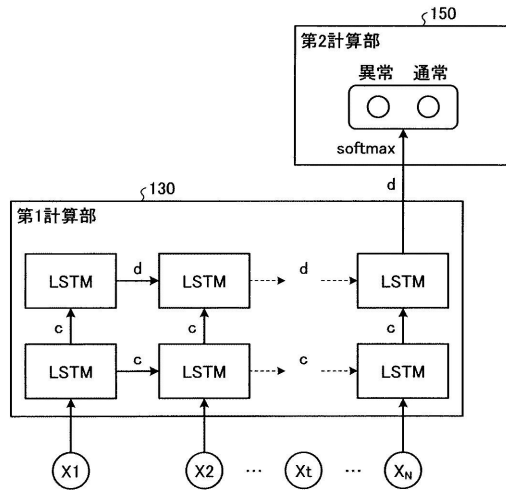
【図 21】

正常な会話の音声データの内部ベクトルの一例を示す図



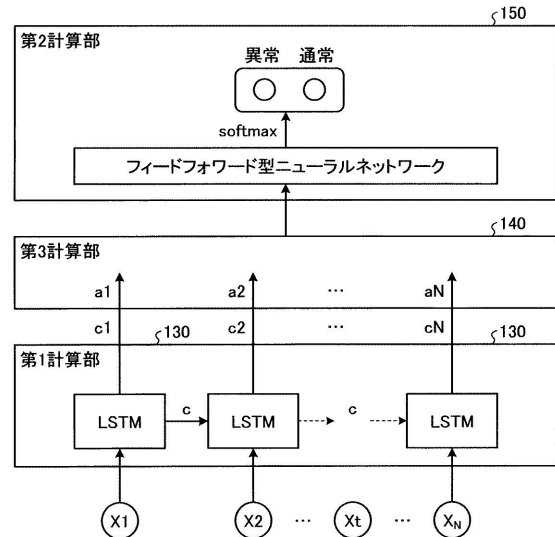
【図 22】

第1計算部および第2計算部の拡張例を説明するための図

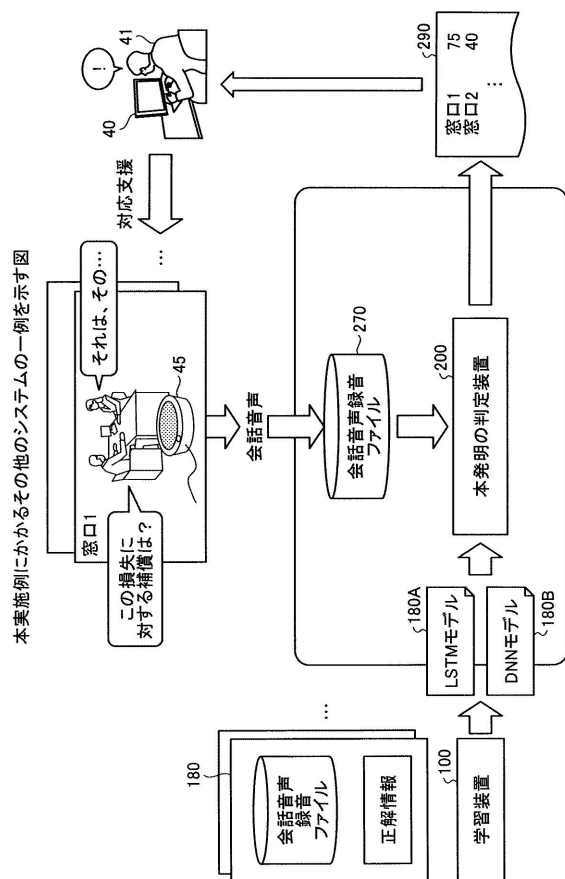


【図 23】

第3計算部のその他の処理を説明するための図



【図 2 4】



【図 25】

ハードウェア構成例を説明する図

