(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2020-129051

（Ｐ２０２０－１２９０５１Ａ）

(43)公開日　令和2年8月27日(2020.8.27)

(51)Int.Cl.			Ｆ Ｉ		テーマコード(参考)
Ｇ１０Ｌ	*15/06*	*(2013.01)*	Ｇ１０Ｌ	15/06	３００Ｙ
Ｇ１０Ｌ	*15/16*	*(2006.01)*	Ｇ１０Ｌ	15/16	
Ｇ１０Ｌ	*15/10*	*(2006.01)*	Ｇ１０Ｌ	15/10	５００Ｎ
Ｇ１０Ｌ	*25/63*	*(2013.01)*	Ｇ１０Ｌ	25/63	

審査請求　未請求　請求項の数8　ＯＬ　（全16頁）

(21)出願番号　特願2019-21332(P2019-21332)

(22)出願日　平成31年2月8日(2019.2.8)

(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町一丁目５番１号

(74)代理人　100121706
弁理士　中尾　直樹

(74)代理人　100128705
弁理士　中村　幸雄

(74)代理人　100147773
弁理士　義村　宗洋

(72)発明者　安藤　厚志
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内

(72)発明者　神山　歩相名
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内

最終頁に続く

(54)【発明の名称】パラ言語情報推定モデル学習装置、パラ言語情報推定装置、およびプログラム

(57)【要約】

【課題】パラ言語情報を特定することが困難である発話をモデル学習に用いた場合であっても、高精度にパラ言語情報を推定する。

【解決手段】音響特徴抽出部１１は、発話から音響特徴を抽出する。逆教師決定部１２は、複数の聴取者が発話ごとに付与したパラ言語情報の判定結果を表すパラ言語情報ラベルに基づいて、その発話のパラ言語情報の正解ではない逆教師を表す逆教師ラベルを決定する。逆教師推定モデル学習部１３は、発話から抽出した音響特徴と逆教師ラベルとに基づいて、入力された音響特徴に対する逆教師の事後確率を出力する逆教師推定モデルを学習する。

【選択図】図２

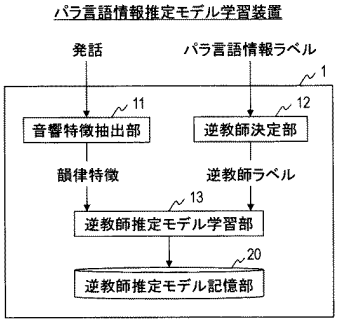

図２

1

【特許請求の範囲】
【請求項１】
複数の聴取者が発話ごとに付与したパラ言語情報の判定結果を表すパラ言語情報ラベルに基づいて、その発話のパラ言語情報の正解ではない逆教師を表す逆教師ラベルを決定する逆教師決定部と、
上記発話から抽出した音響特徴と上記逆教師ラベルとに基づいて、入力された音響特徴に対する逆教師の事後確率を出力する逆教師推定モデルを学習する逆教師推定モデル学習部と、
を含むパラ言語情報推定モデル学習装置。
【請求項２】
請求項１に記載のパラ言語情報推定モデル学習装置であって、
上記パラ言語情報ラベルに基づいて、その発話のパラ言語情報の正解である従来教師を表す従来教師ラベルを決定する従来教師決定部と、
上記発話から抽出した音響特徴と上記従来教師ラベルとに基づいて、入力された音響特徴に対する従来教師の事後確率を出力する従来教師推定モデルを学習する従来教師推定モデル学習部と、
をさらに含むパラ言語情報推定モデル学習装置。
【請求項３】
複数の聴取者が発話ごとに付与したパラ言語情報の判定結果を表すパラ言語情報ラベルに基づいて、その発話のパラ言語情報の正解ではない逆教師を表す逆教師ラベルを決定する逆教師決定部と、
上記パラ言語情報ラベルに基づいて、その発話のパラ言語情報の正解である従来教師を表す従来教師ラベルを決定する従来教師決定部と、
上記発話から抽出した音響特徴と上記逆教師ラベルと上記従来教師ラベルとに基づいてマルチタスク学習を行い、入力された音響特徴に対する逆教師の事後確率と従来教師の事後確率とを出力するマルチタスク推定モデルを学習するマルチタスク推定モデル学習部と、
を含むパラ言語情報推定モデル学習装置。
【請求項４】
請求項３に記載のパラ言語情報推定モデル学習装置であって、
上記マルチタスク推定モデルは、従来教師の事後確率を出力する深層学習に基づくパラ言語情報推定モデルに対して逆教師の事後確率を出力する分岐構造を加えたモデルである、
パラ言語情報推定モデル学習装置。
【請求項５】
請求項１に記載のパラ言語情報推定モデル学習装置が学習した逆教師推定モデルを記憶する逆教師推定モデル記憶部と、
入力発話から抽出した音響特徴を上記逆教師推定モデルに入力して得られる逆教師の事後確率に基づいて上記入力発話のパラ言語情報を推定するパラ言語情報推定部と、
を含むパラ言語情報推定装置。
【請求項６】
請求項５に記載のパラ言語情報推定装置であって、
請求項２に記載のパラ言語情報推定モデル学習装置が学習した従来教師推定モデルを記憶する従来教師推定モデル記憶部をさらに含み、
上記パラ言語情報推定部は、上記音響特徴を上記逆教師推定モデルに入力して得られる逆教師の事後確率と上記音響特徴を上記従来教師推定モデルに入力して得られる従来教師の事後確率との重み付き差に基づいて上記入力発話のパラ言語情報を推定する、
パラ言語情報推定装置。
【請求項７】
請求項３または４に記載のパラ言語情報推定モデル学習装置が学習したマルチタスク推定モデルを記憶するマルチタスク推定モデル記憶部と、
入力発話から抽出した音響特徴を上記マルチタスク推定モデルに入力して得られる従来教師の事後確率に基づいて上記入力発話のパラ言語情報を推定するパラ言語情報推定部と、
を含むパラ言語情報推定装置。
【請求項８】
請求項１から４のいずれかに記載のパラ言語情報推定モデル学習装置もしくは請求項５から７のいずれかに記載のパラ言語情報推定装置としてコンピュータを機能させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声からパラ言語情報を推定する技術に関する。
【背景技術】
【０００２】
音声からパラ言語情報（例えば、感情が喜び・悲しみ・怒り・平静のいずれか）を推定する技術が求められている。パラ言語情報は、音声対話における話し相手の感情を考慮した対話制御（例えば、相手が怒っていれば話題を変える、など）や、音声を用いたメンタルヘルス診断（例えば、毎日の音声を収録し、悲しみや怒り音声の頻度からメンタルヘルス状況を予測する、など）に応用可能である。
【０００３】
従来技術として、機械学習に基づくパラ言語情報推定技術が非特許文献１に示されている。非特許文献１では、図１に示すように、音声から抽出した短時間ごとの音響特徴（例えば、声の高さなど）の時系列情報を入力とし、話者のパラ言語情報を推定する。このとき、再帰型ニューラルネットワーク（Recurrent Neural Network: RN

3

N）と注意機構と呼ばれる機能を組み合わせた深層学習に基づく推定モデルを用いており、音声の部分的な特性に基づいてパラ言語情報を推定することが可能となっている（例えば、発話末尾で声の大きさが急激に低くなったことから、悲しみ感情であると推定することができる）。近年では、非特許文献１のような深層学習に基づくパラ言語情報推定モデルが主流となっている。

【０００４】

なお、従来技術では、ある音声を複数の聴取者が聴取し、聴取者の過半数がその音声に対して特定のパラ言語情報を感じた場合にのみ、その特定のパラ言語情報を正解のパラ言語情報とみなす。従来技術では、この正解のパラ言語情報を推定するように学習を行う。

【先行技術文献】

【非特許文献】

【０００５】

【非特許文献１】S. Mirsamadi, E. Barsoum, and C. Zhang, "Automatic speech emotion recognition using recurrent neural networks with local attention," in Proc. of ICASSP, 2017, pp. 2227-2231.

【発明の概要】

【発明が解決しようとする課題】

【０００６】

しかしながら、従来技術を用いても、パラ言語情報推定の精度が不十分な場合がある。

これは、従来技術では音声に対してただ一つの正解となるパラ言語情報を特定しようと推定モデルを学習するが、正解となるパラ言語情報の特定は人間でも困難な課題であるためである。例えば、パラ言語情報推定のうち感情推定（例えば、喜び・悲しみ・怒り・平静のいずれかを推定する問題）では、従来技術は、ある音声と正解感情との組を学習データとして感情推定モデルを学習する。しかしながら、実際には、正解感情の特定が困難である発話が多く存在する。例えば、３名が聴取した際に、２名が「喜び」、１名が「平静」と判定するような発話が存在する（この場合、従来技術では「喜び」を正解感情とする）。このような発話から正解感情（すなわち「喜び」）に固有の特性を学習することは困難である。この結果、推定モデルを正しく学習することが困難となり、パラ言語情報推定精度が低下するおそれがある。

【０００７】

本発明の目的は、上述のような技術的課題に鑑みて、パラ言語情報を特定することが困難である発話をモデル学習に用いた場合であっても、高精度にパラ言語情報を推定することである。

【課題を解決するための手段】

【０００８】

本発明の第一の態様のパラ言語情報推定モデル学習装置は、複数の聴取者が発話ごとに付与したパラ言語情報の判定結果を表すパラ言語情報ラベルに基づいて、その発

4

話のパラ言語情報の正解ではない逆教師を表す逆教師ラベルを決定する逆教師決定部と、発話から抽出した音響特徴と逆教師ラベルとに基づいて、入力された音響特徴に対する逆教師の事後確率を出力する逆教師推定モデルを学習する逆教師推定モデル学習部と、を含む。

【０００９】

本発明の第二の態様のパラ言語情報推定モデル学習装置は、複数の聴取者が発話ごとに付与したパラ言語情報の判定結果を表すパラ言語情報ラベルに基づいて、その発話のパラ言語情報の正解ではない逆教師を表す逆教師ラベルを決定する逆教師決定部と、パラ言語情報ラベルに基づいて、その発話のパラ言語情報の正解である従来教師を表す従来教師ラベルを決定する従来教師決定部と、発話から抽出した音響特徴と逆教師ラベルと従来教師ラベルとに基づいてマルチタスク学習を行い、入力された音響特徴に対する逆教師の事後確率と従来教師の事後確率とを出力するマルチタスク推定モデルを学習するマルチタスク推定モデル学習部と、を含む。

【００１０】

本発明の第三の態様のパラ言語情報推定装置は、第一の態様のパラ言語情報推定モデル学習装置が学習した逆教師推定モデルを記憶する逆教師推定モデル記憶部と、入力発話から抽出した音響特徴を逆教師推定モデルに入力して得られる逆教師の事後確率に基づいて入力発話のパラ言語情報を推定するパラ言語情報推定部と、を含む。

【００１１】

本発明の第四の態様のパラ言語情報推定装置は、第二の態様のパラ言語情報推定モデル学習装置が学習したマルチタスク推定モデルを記憶するマルチタスク推定モデル記憶部と、入力発話から抽出した音響特徴をマルチタスク推定モデルに入力して得られる従来教師の事後確率に基づいて入力発話のパラ言語情報を推定するパラ言語情報推定部と、を含む。

【発明の効果】

【００１２】

本発明によれば、パラ言語情報を特定することが困難である発話をモデル学習に用いた場合であっても、高精度にパラ言語情報を推定することができる。

【図面の簡単な説明】

【００１３】

【図１】図１は、従来のパラ言語情報推定モデルを説明するための図である。

【図２】図２は、第一実施形態のパラ言語情報推定モデル学習装置の機能構成を例示する図である。

【図３】図３は、第一実施形態のパラ言語情報推定モデル学習方法の処理手順を例示する図である。

【図４】図４は、第一実施形態のパラ言語情報推定装置の機能構成を例示する図である。

【図５】図５は、第一実施形態のパラ言語情報推定方法の処理手順を例示する図である。

5

【図６】図６は、第二実施形態のパラ言語情報推定モデル学習装置の機能構成を例示する図である。
【図７】図７は、第二実施形態のパラ言語情報推定モデル学習方法の処理手順を例示する図である。
【図８】図８は、第二実施形態のパラ言語情報推定装置の機能構成を例示する図である。
【図９】図９は、第二実施形態のパラ言語情報推定方法の処理手順を例示する図である。
【図１０】図１０は、第三実施形態のパラ言語情報推定モデル学習装置の機能構成を例示する図である。
【図１１】図１１は、第三実施形態のパラ言語情報推定モデル学習方法の処理手順を例示する図である。
【図１２】図１２は、第三実施形態のパラ言語情報推定装置の機能構成を例示する図である。
【図１３】図１３は、第三実施形態のパラ言語情報推定方法の処理手順を例示する図である。
【図１４】図１４は、第三実施形態のパラ言語情報推定モデルを説明するための図である。
【発明を実施するための形態】
【００１４】
文中で使用する記号「^」は、本来直後の文字の真上に記載されるべきものであるが、テキスト記法の制限により、当該文字の直前に記載する。数式中においてはこれらの記号は本来の位置、すなわち文字の真上に記述している。例えば、「^c」は数式中では次式で表される。
【００１５】
【数１】

$$\hat{c}$$

【００１６】
以下、本発明の実施の形態について詳細に説明する。なお、図面中において同じ機能を有する構成部には同じ番号を付し、重複説明を省略する。
【００１７】
［発明のポイント］
本発明のポイントは、あえて「絶対に正解ではないパラ言語情報」を推定することで、正解となるパラ言語情報の特定に貢献する点にある。人間にとってパラ言語情報は、ただ一つの正解の特定は困難である一方で、絶対に正解ではないものの推定は容易であると考えられる。例えば、人間がある音声を聴取した際に、喜びか平静かを特定することは困難な場合があるが、そのような音声でも「怒りではない」「悲しみでもない」と判断することは可能であることが多い。このことから、絶対に正解ではないパラ言語情報の推定は正解のパラ言語情報の特定に比べて容易である可能性があり、高い精度で絶対に正解ではないパラ言語情報の推定を行うことができると考えられる。また、消去法のような枠組みを用いることで、絶対に正解ではないパラ言語情報がわかることは、ただ一つの正解となるパラ言語情報の特定にも貢献できる。以降では、ただ一つの正解となるパラ言語情報を「従

6

来教師」、絶対に正解ではないパラ言語情報を「逆教師」と呼ぶ。
【００１８】
上記の発明のポイントを実現するために、後述する実施形態は下記のように構成される。
【００１９】
１．複数名の聴取者によるパラ言語情報の判定結果に基づき、逆教師を決定する。本発明では、逆教師とは、推定対象となるパラ言語情報のうち、一定以下（例えば１割以下）の聴取者が判定したパラ言語情報を指すものとする。例えば、喜び・悲しみ・怒り・平静の４クラスの感情推定において、ある音声に対して３名の聴取者が「喜び」「喜び」「平静」と判定した場合、その音声の逆教師は「悲しみ」「怒り」の２クラスを指す。
【００２０】
２．逆教師の推定モデルを学習する。この推定モデルは、入力特徴量と推定モデル構造が従来技術と同様であるが、最終的な推定部がマルチラベル分類（一つの音声が複数のクラスに同時に分類され得る）の構造を持つモデルを利用することで実現される。
【００２１】
３．逆教師の推定モデル単独、または逆教師の推定モデルと従来教師の推定モデルの両方を用いてパラ言語情報を推定する。逆教師の推定モデルを単独で用いる場合であれば、逆教師の推定モデルによる逆教師推定を行い、その出力確率が最も小さい（すなわち、絶対に正解ではないパラ言語情報の確率が最も小さい）クラスを正解のパラ言語情報推定結果とみなす。逆教師の推定モデルと従来教師の推定モデルの両方を用いる場合であれば、従来教師の推定モデルの出力確率から、逆教師の推定モデルの出力確率を減算した値（すなわち、正解であるパラ言語情報の確率から正解でないパラ言語情報の確率を減算した値）が最も大きいクラスを正解のパラ言語情報推定結果とみなす。
【００２２】
［第一実施形態］
第一実施形態では、逆教師の推定モデルを単独で用いてパラ言語情報を推定する。
【００２３】
＜パラ言語情報推定モデル学習装置１＞
第一実施形態のパラ言語情報推定モデル学習装置は、複数の発話と、発話ごとに複数の聴取者が付与したパラ言語情報の判定結果を表すパラ言語情報ラベルとからなる学習データから、逆教師推定モデルを学習する。第一実施形態のパラ言語情報推定モデル学習装置１は、図２に例示するように、音響特徴抽出部１１、逆教師決定部１２、逆教師推定モデル学習部１３、および逆教師推定モデル記憶部１０を備える。このパラ言語情報推定モデル学習装置１が、図３に例示する各ステップの処理を行うことにより第一実施形態のパラ言語情報推定モデル学習

7

方法が実現される。

【００２４】

パラ言語情報推定モデル学習装置１は、例えば、中央演算処理装置（CPU: Central Processing Unit）、主記憶装置（RAM: Random Access Memory）などを有する公知又は専用のコンピュータに特別なプログラムが読み込まれて構成された特別な装置である。パラ言語情報推定モデル学習装置１は、例えば、中央演算処理装置の制御のもとで各処理を実行する。パラ言語情報推定モデル学習装置１に入力されたデータや各処理で得られたデータは、例えば、主記憶装置に格納され、主記憶装置に格納されたデータは必要に応じて中央演算処理装置へ読み出されて他の処理に利用される。パラ言語情報推定モデル学習装置１の各処理部は、少なくとも一部が集積回路等のハードウェアによって構成されていてもよい。パラ言語情報推定モデル学習装置１が備える各記憶部は、例えば、RAM（Random Access Memory）などの主記憶装置、ハードディスクや光ディスクもしくはフラッシュメモリ（Flash Memory）のような半導体メモリ素子により構成される補助記憶装置、またはリレーショナルデータベースやキーバリューストアなどのミドルウェアにより構成することができる。

【００２５】

ステップＳ１１において、音響特徴抽出部１１は、学習データの発話から韻律特徴を抽出する。韻律特徴は、基本周波数、短時間パワー、メル周波数ケプストラム係数（Mel-frequency Cepstral Coefficients: MFCC）、ゼロ交差率、Harmonics-to-Noise-Ratio (HNR)、メルフィルタバンク出力、のいずれか一つ以上の特徴量を含むベクトルである。また、これらの時間ごと（フレームごと）の系列ベクトルであってもよいし、これらの一定時間ごとまたは発話全体の統計量（平均、分散、最大値、最小値、勾配など）のベクトルであってもよい。音響特徴抽出部１１は、抽出した韻律特徴を逆教師推定モデル学習部１３へ出力する。

【００２６】

ステップＳ１２において、逆教師決定部１２は、学習データのパラ言語情報ラベルから逆教師ラベルを決定する。逆教師とは、推定対象となるパラ言語情報のうち、予め定めた閾値（以下、「逆教師閾値」と呼ぶ）以下（例えば、１割以下）の聴取者が判定したパラ言語情報を指すものとする。逆教師ラベルは、逆教師のパラ言語情報クラスが１、それ以外は０となるベクトルを指す。すなわち、従来教師のようにいずれか一つのパラ言語情報クラスが１、それ以外が０となるベクトルではなく、少なくとも一つ以上のパラ言語情報クラスが１となるベクトルが逆教師ラベルである。例えば、喜び・悲しみ・怒り・平静の４クラスの感情推定において、逆教師閾値を0.1とし、３名の聴取者がある音声に対して「喜び」「喜び」「平静」と判定した場合、その音声の逆教師ラベル

8

は「悲しみ」「怒り」の２クラスが１、「喜び」「平静」の２クラスが０となる４次元のベクトルを指す。

【００２７】

逆教師ラベルは、具体的には以下のように表される。

【００２８】

【数２】

$$t^* = \begin{bmatrix} t_1^* \\ \vdots \\ t_K^* \end{bmatrix},$$

$$t_k^* = \begin{cases} 1 & if \ \dfrac{1}{N}\sum_{n=1}^{N} h_k^n \le \beta \\ 0 & otherwise \end{cases}$$

【００２９】

ここで、h_k^nはn番目の聴取者がk番目のパラ言語情報クラスを感じたか否か（感じた場合は１、感じなかった場合は０）を表す。Kはパラ言語情報クラスの総数である。Nは聴取者の総数である。βは０以上１以下の逆教師閾値である。

【００３０】

逆教師ラベルは、いずれの聴取者も判定しなかったパラ言語情報クラスを逆教師としてもよい。これは、逆教師閾値βを０に設定した場合に相当する。

【００３１】

逆教師決定部１２は、決定した逆教師ラベルを逆教師推定モデル学習部１３へ出力する。

【００３２】

ステップＳ１３において、逆教師推定モデル学習部１３は、音響特徴抽出部１１が出力する韻律特徴と逆教師決定部１２が出力する逆教師ラベルとに基づいて、逆教師推定モデルを学習する。推定モデルにはマルチラベル分類問題（一つの音声が複数のクラスに同時に分類されうる分類問題）を扱うことができるモデルを用いるものとする。これは一つの音声に対して複数のクラスに逆教師が表れることがあるためである。推定モデルは例えば従来技術のような深層学習に基づくモデルであってもよいし、多クラスロジスティック回帰であってもよいが、出力が確率値（あるパラ言語情報クラスが１である確率）として表現可能であるものとする。逆教師推定モデル学習部１３は、学習した逆教師推定モデルを逆教師推定モデル記憶部２０へ記憶する。

【００３３】

＜パラ言語情報推定装置２＞

第一実施形態のパラ言語情報推定装置は、学習済みの逆教師推定モデルを用いて、入力された発話のパラ言語情報を推定する。第一実施形態のパラ言語情報推定装置２は、図４に例示するように、音響特徴抽出部１１、逆教

9

師推定モデル記憶部２０、およびパラ言語情報推定部２１を備える。このパラ言語情報推定装置２が、図５に例示する各ステップの処理を行うことにより第一実施形態のパラ言語情報推定方法が実現される。

【００３４】

パラ言語情報推定装置２は、例えば、中央演算処理装置（CPU: Central Processing Unit）、主記憶装置（RAM: Random Access Memory）などを有する公知又は専用のコンピュータに特別なプログラムが読み込まれて構成された特別な装置である。パラ言語情報推定装置２は、例えば、中央演算処理装置の制御のもとで各処理を実行する。パラ言語情報推定装置２に入力されたデータや各処理で得られたデータは、例えば、主記憶装置に格納され、主記憶装置に格納されたデータは必要に応じて中央演算処理装置へ読み出されて他の処理に利用される。パラ言語情報推定装置２の各処理部は、少なくとも一部が集積回路等のハードウェアによって構成されていてもよい。パラ言語情報推定装置２が備える各記憶部は、例えば、RAM（Random Access Memory）などの主記憶装置、ハードディスクや光ディスクもしくはフラッシュメモリ（Flash Memory）のような半導体メモリ素子により構成される補助記憶装置、またはリレーショナルデータベースやキーバリューストアなどのミドルウェアにより構成することができる。

【００３５】

逆教師推定モデル記憶部２０には、パラ言語情報推定モデル学習装置１が学習した逆教師推定モデルが記憶されている。

【００３６】

ステップＳ１１において、音響特徴抽出部１１は、入力された発話から韻律特徴を抽出する。韻律特徴の抽出は、パラ言語情報推定モデル学習装置１と同様に行えばよい。音響特徴抽出部１１は、抽出した韻律特徴をパラ言語情報推定部２１へ出力する。

【００３７】

ステップＳ２１において、パラ言語情報推定部２１は、逆教師推定モデル記憶部２０に記憶されている逆教師推定モデルに基づいて、音響特徴抽出部１１が出力した韻律特徴からパラ言語情報を推定する。推定においては、ある韻律特徴に対し、逆教師推定モデルの出力が最も低いクラスをパラ言語情報推定結果とみなす。これは、逆教師の可能性が最も低いパラ言語情報、すなわち「絶対に正解ではないパラ言語情報」ではないと考えられるパラ言語情報を選択することに相当する。パラ言語情報推定部２１は、パラ言語情報の推定結果をパラ言語情報推定装置２の出力とする。

【００３８】

［第二実施形態］

第二実施形態では、逆教師の推定モデルに加えて、従来教師の推定モデルを用いてパラ言語情報を推定する。このとき、各推定モデルの出力結果の重み付け差によりパラ言語情報の推定を行う。「あるパラ言語情報が正解である確率」と「あるパラ言語情報が正解でない確率」の両方を考慮してパラ言語情報推定を行うことに相当する。その結果、どちらかの確率のみを考慮する場合（すなわち、従来技術や第一実施形態のそれぞれ）に比べてパラ言語情報の推定精度が向上する。

【００３９】

＜パラ言語情報推定モデル学習装置３＞

第二実施形態のパラ言語情報推定モデル学習装置は、第一実施形態と同様の学習データから逆教師推定モデルと従来教師推定モデルを学習する。第二実施形態のパラ言語情報推定モデル学習装置３は、図６に例示するように、第一実施形態の音響特徴抽出部１１、逆教師決定部１２、逆教師推定モデル学習部１３、および逆教師推定モデル記憶部２０に加えて、従来教師決定部３１、従来教師推定モデル学習部３２、および従来教師推定モデル記憶部４０をさらに備える。このパラ言語情報推定モデル学習装置３が、図７に例示する各ステップの処理を行うことにより第二実施形態のパラ言語情報推定モデル学習方法が実現される。

【００４０】

以下、第一実施形態のパラ言語情報推定モデル学習装置１との相違点を中心に、第二実施形態のパラ言語情報推定モデル学習装置３を説明する。

【００４１】

ステップＳ３１において、従来教師決定部３１は、学習データのパラ言語情報ラベルから従来教師ラベルを決定する。従来教師ラベルは、従来技術と同様に、ある音声に対して全聴取者のうち過半数が同じパラ言語情報を判定した場合に、そのパラ言語情報クラスを１、それ以外は０としたベクトルである。過半数が同じパラ言語情報を判定しなかった場合、その音声は正解なしとしてモデル学習には利用しない。例えば、喜び・悲しみ・怒り・平静の４クラスの感情推定において、ある音声に対して３名の聴取者が「喜び」「喜び」「平静」と判定した場合、その音声の従来教師ラベルは「喜び」クラスが１、残りの「悲しみ」「怒り」「平静」の３クラスが０となる４次元のベクトルを指す。従来教師決定部３１は、決定した従来教師ラベルを従来教師推定モデル学習部３２へ出力する。

【００４２】

ステップＳ３２において、従来教師推定モデル学習部３２は、音響特徴抽出部１１が出力する韻律特徴と従来教師決定部３１が出力する従来教師ラベルとに基づいて、従来教師推定モデルを学習する。推定モデルには多クラス分類問題（一つの音声から一つのクラスを分類する分類問題）を扱うことができるモデルを用いるものとする。推定モデルは例えば従来技術のような深層学習に基づくモデルであってもよいし、多クラスロジスティック回

11

帰であってもよいが、出力が確率値（あるパラ言語情報クラスが１である確率）として表現可能であるものとする。従来教師推定モデル学習部３２は、学習した従来教師推定モデルを従来教師推定モデル記憶部４０へ記憶する。

【００４３】

＜パラ言語情報推定装置４＞

第二実施形態のパラ言語情報推定装置は、学習済みの逆教師推定モデルと従来教師推定モデルの両方を用いて、入力された発話のパラ言語情報を推定する。第二実施形態のパラ言語情報推定装置４は、図８に例示するように、第一実施形態の音響特徴抽出部１１および逆教師推定モデル記憶部２０に加えて、従来教師推定モデル記憶部４０およびパラ言語情報推定部４１をさらに備える。このパラ言語情報推定装置４が、図９に例示する各ステップの処理を行うことにより第二実施形態のパラ言語情報推定方法が実現される。

【００４４】

以下、第一実施形態のパラ言語情報推定装置２との相違点を中心に、第二実施形態のパラ言語情報推定装置４を説明する。

【００４５】

従来教師推定モデル記憶部４０には、パラ言語情報推定モデル学習装置３が学習した従来教師推定モデルが記憶されている。

【００４６】

ステップＳ４１において、パラ言語情報推定部４１は、逆教師推定モデル記憶部２０に記憶されている逆教師推定モデルと従来教師推定モデル記憶部４０に記憶されている従来教師推定モデルの両方に基づいて、音響特徴抽出部１１が出力した韻律特徴からパラ言語情報を推定する。推定においては、ある韻律特徴に対する従来教師推定モデルの出力と逆教師推定モデルの出力との重み付け差により決定する。これは、「あるパラ言語情報が正解である確率」と「あるパラ言語情報が正解でない確率」の両方を考慮してパラ言語情報推定を行うことに相当する。

【００４７】

パラ言語情報の推定は、具体的には以下のように表される。

【００４８】

【数３】

$$\hat{c}_k = \arg\max_{c_k}\left((1-\alpha)p(c_k) - \alpha q(c_k)\right)$$

【００４９】

ここで、$\hat{c}_k$はパラ言語情報の推定結果を表す。c_kはk番目のパラ言語情報クラスを表す。

$p(c_k)$はk番目のパラ言語情報クラスが正解である確率を表し、従来教師推定モデルの出力である。$q(c_k)$はk番目

12

のパラ言語情報クラスが正解でない確率を表し、逆教師推定モデルの出力である。αは推定重みを表す。

【００５０】

推定重みαは０から１までの連続値のいずれかの値とする。推定重みが０に近いほど「あるパラ言語情報が正解である確率」を重視し、１に近いほど「あるパラ言語情報が正解でない確率」を重視した推定を行うこととなる。例えば、推定重みは0.3とする。

【００５１】

［第二実施形態の変形例］

第二実施形態では、逆教師推定モデル学習部と従来教師推定モデル学習部を完全に別々の処理として構成する例を説明したが、一方で学習した推定モデルを他方の推定モデル初期値として利用しながら各推定モデルの学習を行うことも可能である。すなわち、従来教師推定モデル学習部３２に逆教師推定モデル記憶部２０で記憶されている逆教師推定モデルを入力し、従来教師推定モデル学習部３２では逆教師推定モデルを従来教師推定モデルの初期値として、音響特徴抽出部１１が出力する韻律特徴と従来教師決定部３１が出力する従来教師ラベルとに基づいて、従来教師推定モデルを学習する。または、逆教師推定モデル学習部１３に従来教師推定モデル記憶部４０で記憶されている従来教師推定モデルを入力し、逆教師推定モデル学習部１３では従来教師推定モデルを逆教師推定モデルの初期値として、音響特徴抽出部１１が出力する韻律特徴と逆教師決定部１２が出力する逆教師ラベルとに基づいて、逆教師推定モデルを学習する。従来教師推定モデルと逆教師推定モデルはそれぞれ韻律特徴と従来教師または韻律特徴と逆教師との関連性を学習しており、一方で学習された推定基準は他方でも利用可能であると考えられるため、これらの処理を加えることでパラ言語情報推定精度がさらに向上する可能性がある。

【００５２】

［第三実施形態］

第三実施形態では、従来教師の推定と逆教師の推定とを同時に行うマルチタスク推定モデルを用いてパラ言語情報を推定する。このとき、モデル学習において、従来教師の推定と逆教師の推定をマルチタスク学習として同時に学習する。マルチタスク学習は、異なる課題を単一のモデルで解くことにより、課題間で共通する知識を獲得でき、各課題の推定精度が向上することが知られている（下記参考文献１参照）。第三実施形態は、第二実施形態と同様に従来教師と逆教師の両方を用いたパラ言語情報推定であるが、マルチタスク学習として学習することで推定モデル自体を改善することができるため、推定精度がより向上する。

〔参考文献１〕R. Caruana, "Multitask Learning", Machine Learning, vol. 28, pp.41-75, 1997.

【００５３】

＜パラ言語情報推定モデル学習装置５＞

13

第三実施形態のパラ言語情報推定モデル学習装置は、第一実施形態と同様の学習データからマルチタスク推定モデルを学習する。第三実施形態のパラ言語情報推定モデル学習装置５は、図１０に例示するように、第一実施形態の音響特徴抽出部１１および逆教師決定部１２と、第二実施形態の従来教師決定部３１とに加えて、マルチタスク推定モデル学習部５１およびマルチタスク推定モデル記憶部６０をさらに備える。このパラ言語情報推定モデル学習装置５が、図１１に例示する各ステップの処理を行うことにより第三実施形態のパラ言語情報推定モデル学習方法が実現される。

【００５４】

以下、第一実施形態のパラ言語情報推定モデル学習装置１および第二実施形態のパラ言語情報推定モデル学習装置３との相違点を中心に、第三実施形態のパラ言語情報推定モデル学習装置５を説明する。

【００５５】

ステップＳ５１において、マルチタスク推定モデル学習部５１は、音響特徴抽出部１１が出力する韻律特徴と逆教師決定部１２が出力する逆教師ラベルと従来教師決定部３１が出力する従来教師ラベルとを用いてマルチタスク学習を行い、マルチタスク推定モデルを学習する。マルチタスク学習では、一般的にニューラルネットワークに基づく推定モデルが用いられるため、本実施形態での推定モデルはニューラルネットワークに基づく推定モデルとする。例えば、図１０に示すように、従来技術の深層学習に基づく推定モデルに対して、逆教師の推定を行う分岐構造を加えた推定モデルとする。マルチタスク推定モデル学習部５１は、学習したマルチタスク推定モデルをマルチタスク推定モデル記憶部６０へ記憶する。

【００５６】

＜パラ言語情報推定装置６＞

第三実施形態のパラ言語情報推定装置は、学習済みのマルチタスク推定モデルを用いて、入力された発話のパラ言語情報を推定する。第三実施形態のパラ言語情報推定装置６は、図１２に例示するように、第一実施形態の音響特徴抽出部１１に加えて、マルチタスク推定モデル記憶部６０およびパラ言語情報推定部６１をさらに備える。このパラ言語情報推定装置６が、図１３に例示する各ステップの処理を行うことにより第三実施形態のパラ言語情報推定方法が実現される。

【００５７】

以下、第一実施形態のパラ言語情報推定装置２および第二実施形態のパラ言語情報推定装置４との相違点を中心に、第三実施形態のパラ言語情報推定装置６を説明する。

【００５８】

マルチタスク推定モデル記憶部６０には、パラ言語情報推定モデル学習装置５が学習したマルチタスク推定モデルが記憶されている。

14

【００５９】

ステップＳ６１において、パラ言語情報推定部６１は、マルチタスク推定モデル記憶部６０に記憶されているマルチタスク推定モデルに基づいて、音響特徴抽出部１１が出力した韻律特徴からパラ言語情報を推定する。推定においては、ある韻律特徴に対し、従来教師の推定出力が最も高いクラスをパラ言語情報推定結果とみなす。推定モデルの学習においてマルチタスク学習を用いているため、逆教師の影響を考慮した（すなわち、逆教師を間違えないようにしつつ従来教師を推定している）パラ言語情報推定を行うことができ、パラ言語情報推定精度が向上する。

【００６０】

［変形例］

上述の実施形態では、パラ言語情報推定モデル学習装置とパラ言語情報推定装置とを別々の装置として構成する例を説明したが、パラ言語情報推定モデル学習する機能と学習済みのパラ言語情報推定モデルを用いてパラ言語情報を推定する機能とを兼ね備えた１台のパラ言語情報推定装置として構成することも可能である。すなわち、第一実施形態の変形例のパラ言語情報推定装置は、音響特徴抽出部１１、逆教師決定部１２、逆教師推定モデル学習部１３、逆教師推定モデル記憶部２０、およびパラ言語情報推定部２１を備える。

また、第二実施形態の変形例のパラ言語情報推定装置は、音響特徴抽出部１１、逆教師決定部１２、逆教師推定モデル学習部１３、および逆教師推定モデル記憶部２０に加えて、従来教師決定部３１、従来教師推定モデル学習部３２、従来教師推定モデル記憶部４０、およびパラ言語情報推定部４１をさらに備える。そして、第三実施形態の変形例のパラ言語情報推定装置は、音響特徴抽出部１１、逆教師決定部１２、および従来教師決定部３１に加えて、マルチタスク推定モデル学習部５１、マルチタスク推定モデル記憶部６０、およびパラ言語情報推定部６１をさらに備える。

【００６１】

以上、本発明の実施の形態について説明したが、具体的な構成は、これらの実施の形態に限られるものではなく、本発明の趣旨を逸脱しない範囲で適宜設計の変更等があっても、本発明に含まれることはいうまでもない。実施の形態において説明した各種の処理は、記載の順に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。

【００６２】

［プログラム、記録媒体］

上記実施形態で説明した各装置における各種の処理機能をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することに

15

より、上記各装置における各種の処理機能がコンピュータ上で実現される。

【００６３】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等どのようなものでもよい。

【００６４】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

【００６５】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記憶装置に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP（Application Service Provider）型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの（コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等）を含むものとする。

16

【００６６】

また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、本装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

【符号の説明】

【００６７】

１，３，５　パラ言語情報推定モデル学習装置
１１　音響特徴抽出部
１２　逆教師決定部
１３　逆教師推定モデル学習部
２０　逆教師推定モデル記憶部
３１　従来教師決定部
３２　従来教師推定モデル学習部
４０　従来教師推定モデル記憶部
５１　マルチタスク推定モデル学習部
６０　マルチタスク推定モデル記憶部
２，４，６　パラ言語情報推定装置
２１，４１，６１　パラ言語情報推定部

【図１】

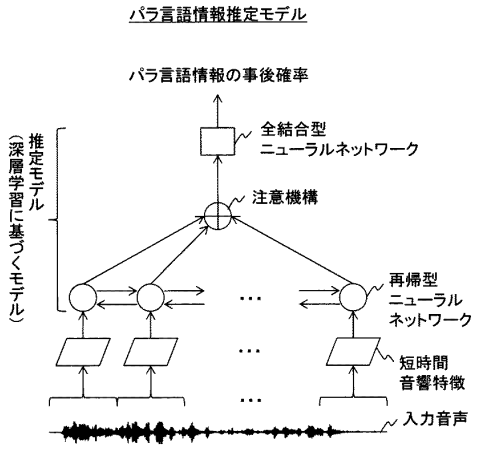

図1

【図２】

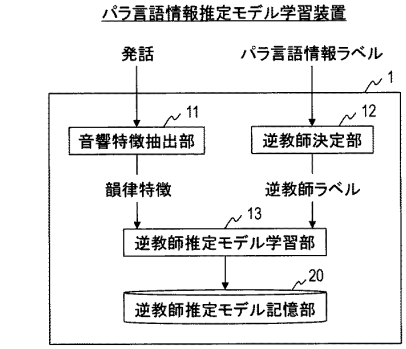

図2

【図３】

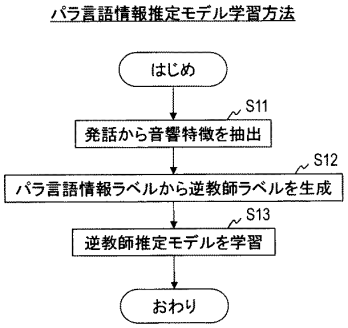

図3

【図４】

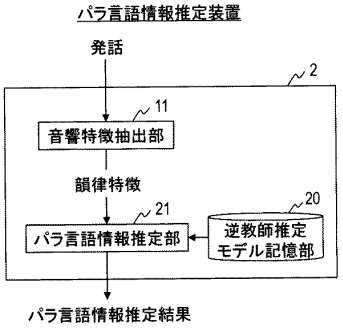

図4

【図５】

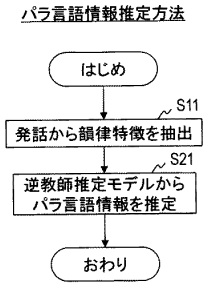

図5

【図６】

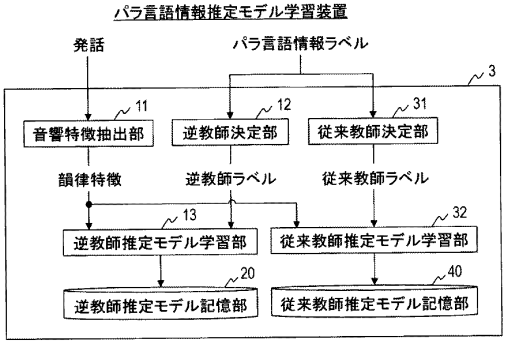

図6

【図７】

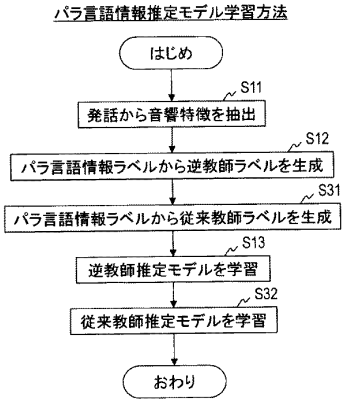

図7

【図８】

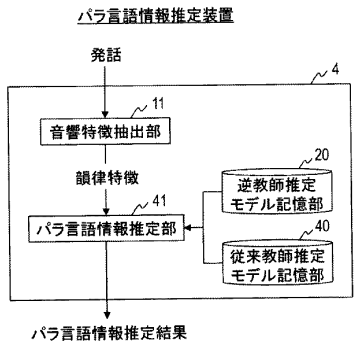

図8

【図９】

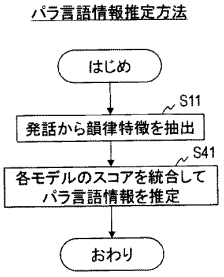

図9

【図１０】

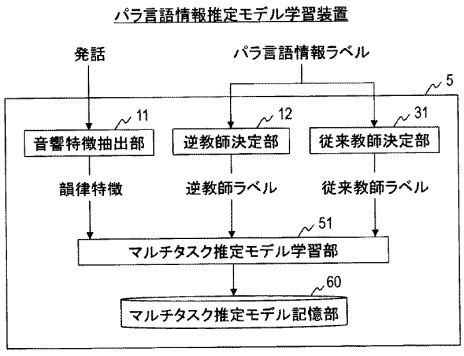

図10

【図１１】

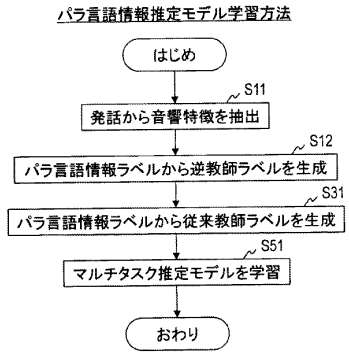

図11

【図１２】

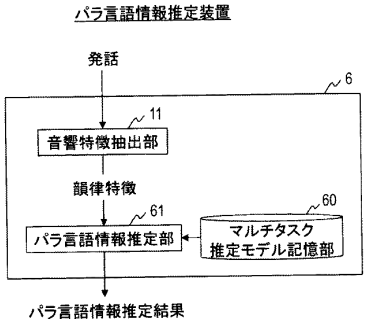

図12

【図１３】

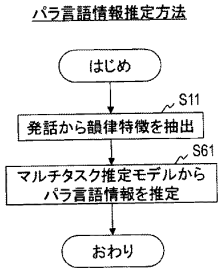

図13

【図１４】

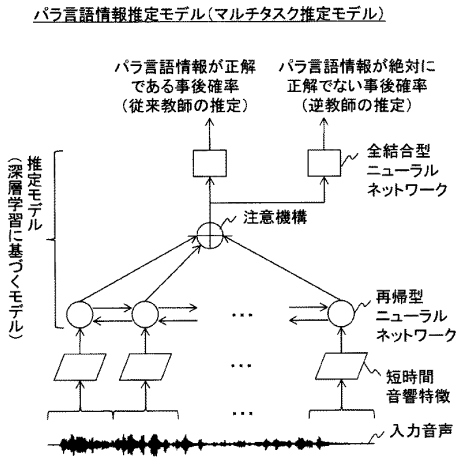

図14

フロントページの続き

(72)発明者　小橋川　哲
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内