

(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2015-200913

(P2015-200913A)

(43) 公開日 平成27年11月12日(2015.11.12)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 O L 17/14 (2013.01)	G 1 O L 17/14	
G 1 O L 17/00 (2013.01)	G 1 O L 17/00	2 O O D
G 1 O L 15/00 (2013.01)	G 1 O L 15/00	2 O O U

審査請求 有 請求項の数 9 O L (全 13 頁)

(21) 出願番号	特願2015-138079 (P2015-138079)	(71) 出願人	000003078
(22) 出願日	平成27年7月9日 (2015.7.9)		株式会社東芝
(62) 分割の表示	特願2011-166071 (P2011-166071) の分割		東京都港区芝浦一丁目1番1号
原出願日	平成23年7月28日 (2011.7.28)	(74) 代理人	100111121 弁理士 原 拓実
		(74) 代理人	100125667 弁理士 小林 幹雄
		(74) 代理人	100138601 弁理士 山下 正成
		(74) 代理人	100151323 弁理士 泉 剛司
		(74) 代理人	100149629 弁理士 柘 周作
		(74) 代理人	100200229 弁理士 矢作 徹夫

最終頁に続く

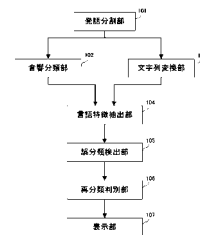
(54) 【発明の名称】 話者分類装置、話者分類方法および話者分類プログラム

(57) 【要約】

【課題】発話に含まれる言語的な特徴を利用して、音響的な特徴による誤分類を検出する話者分類装置を提供することである。

【解決手段】実施形態の話者分類装置は、音響分類手段と、言語特徴抽出手段と、誤分類検出手段とを備える。音響分類手段は、入力された音響信号に含まれる発話を、前記音響信号から抽出した音響的な特徴を利用して話者毎に分類し分類結果を取得する。言語特徴抽出手段は、前記発話の内容を表す文字列を取得し、この文字列を利用して前記分類結果に含まれる話者の言語的な特徴を抽出する。誤分類検出手段は、前記発話の内容を表す文字列が、前記分類結果においてこの発話が分類された話者の前記言語的な特徴に適合するか否かを判別し、適合しない場合は、この発話は前記音響分類手段によって誤分類された発話であると判別する。

【選択図】 図 1



【特許請求の範囲】**【請求項 1】**

入力された音響信号に含まれる発話を、前記音響信号から抽出した音響的な特徴を利用して話者毎に分類し分類結果を取得する音響分類手段と、
前記発話の内容を表す文字列を取得し、この文字列を利用して前記分類結果に含まれる話者の言語的な特徴を抽出する言語特徴抽出手段と、
前記発話の内容を表す文字列が、前記分類結果においてこの発話が分類された話者の前記言語的な特徴に適合するか否かを判別し、適合しない場合は、この発話は前記音響分類手段によって誤分類された発話であると判別する誤分類検出手段と、
前記誤分類された発話が選択された場合、前記誤分類された発話の直前の発話の話者以外の話者を、再分類結果の候補であると判別する再分類判別手段と、
を備える話者分類装置。

10

【請求項 2】

前記言語特徴抽出手段が、予め記憶した言語に関するルールと各話者の前記発話の内容を表す文字列とが適合するか否かを判別し、適合する発話の数が所定閾値を超えるルールを前記話者の言語的な特徴とする請求項 1 記載の話者分類装置。

【請求項 3】

前記入力された音響信号に含まれる発話を音声認識して文字列に変換する文字列変換手段を更に備え、
前記発話の内容を表す文字列が、前記文字列変換手段で変換された文字列である請求項 1 乃至請求項 2 記載の話者分類装置。

20

【請求項 4】

前記音響分類手段の分類結果もしくは前記再分類判別手段での判別結果を表示する表示手段を更に備える請求項 1 記載の話者分類装置。

【請求項 5】

入力された音響信号に含まれる発話を、前記音響信号から抽出した音響的な特徴を利用して話者毎に分類し分類結果を取得する音響分類工程と、
前記発話の内容を表す文字列を取得し、この文字列を利用して前記分類結果に含まれる話者の言語的な特徴を抽出する言語特徴抽出工程と、
前記発話の内容を表す文字列が、前記分類結果においてこの発話が分類された話者の前記言語的な特徴に適合するか否かを判別し、適合しない場合は、この発話は前記音響分類工程によって誤分類された発話であると判別する誤分類検出工程と、
前記誤分類された発話が選択された場合、前記誤分類された発話の直前の発話の話者以外の話者を、再分類結果の候補であると判別する再分類判別工程と、
を備える話者分類方法。

30

【請求項 6】

話者分類装置に、
入力された音響信号に含まれる発話を、前記音響信号から抽出した音響的な特徴を利用して話者毎に分類し分類結果を取得する音響分類工程と、
前記発話の内容を表す文字列を取得し、この文字列を利用して前記分類結果に含まれる話者の言語的な特徴を抽出する言語特徴抽出工程と、
前記発話の内容を表す文字列が、前記分類結果においてこの発話が分類された話者の前記言語的な特徴に適合するか否かを判別し、適合しない場合は、この発話は前記音響分類工程によって誤分類された発話であると判別する誤分類検出工程と、
前記誤分類された発話が選択された場合、前記誤分類された発話の直前の発話の話者以外の話者を、再分類結果の候補であると判別する再分類判別工程と、
を実現させるための話者分類プログラム。

40

【請求項 7】

入力された発話を、話者毎に分類し分類結果を取得する分類手段と、
前記発話の内容を表す文字列が、前記分類結果においてこの発話が分類された話者に適合

50

するか否かを判別し、適合しない場合は、この発話は前記分類手段によって誤分類された発話であると判別する誤分類検出手段と、
前記誤分類された発話を選択された場合、前記誤分類された発話の直前の発話の話者以外の話者を、再分類結果の候補であると判別する再分類判別手段と、
を備える話者分類装置。

【請求項 8】

入力された発話を、話者毎に分類し分類結果を取得する分類工程と、
前記発話の内容を表す文字列が、前記分類結果においてこの発話が分類された話者に適合するか否かを判別し、適合しない場合は、この発話は前記分類工程によって誤分類された発話であると判別する誤分類検出工程と、
前記誤分類された発話を選択された場合、前記誤分類された発話の直前の発話の話者以外の話者を、再分類結果の候補であると判別する再分類判別工程と、
を備える話者分類方法。

10

【請求項 9】

話者分類装置に、
入力された発話を、話者毎に分類し分類結果を取得する分類工程と、
前記発話の内容を表す文字列が、前記分類結果においてこの発話が分類された話者に適合するか否かを判別し、適合しない場合は、この発話は前記分類工程によって誤分類された発話であると判別する誤分類検出工程と、
前記誤分類された発話を選択された場合、前記誤分類された発話の直前の発話の話者以外の話者を、再分類結果の候補であると判別する再分類判別工程と、
を実現させるための話者分類プログラム。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、話者分類装置、話者分類方法および話者分類プログラムに関する。

【背景技術】

【0002】

会議の議事録作成を支援するために、収録された会議音声に含まれる発話を話者毎に分類することが要求される。会議音声から抽出した音響的な特徴（音響特徴量）を利用して、発話を話者毎に分類する技術は多数報告されている。例えば、各発話の音響特徴量と予め学習した多数の話者モデルとの類似度を計算し、この類似度のパターンを基に各発話を話者毎にクラスタリングする技術がある。

30

【0003】

しかしながら、上述したような音響的な特徴のみを利用した方法では、背景雑音があるなど会議音声の音質が劣化した場合に、各発話を正確に分類することができなかった。

【先行技術文献】

【特許文献】

【0004】

40

【特許文献 1】特開 2010 - 60850 号公報

【非特許文献】

【0005】

【非特許文献 1】秋田、河原、「多数話者モデルを用いた討論音声の教師なし話者インデキシング」、電子情報通信学会論文誌、2004 年 2 月 1 日、D-II、情報・システム、II-パターン処理 J87-D-II(2)、p.495-503.

【発明の概要】

【発明が解決しようとする課題】

【0006】

発明が解決しようとする課題は、発話に含まれる言語的な特徴を利用して、音響的な特

50

徴による誤分類を検出する話者分類装置を提供することである。

【課題を解決するための手段】

【0007】

実施形態の話者分類装置は、実施形態の話者分類装置は、音響分類手段と、言語特徴抽出手段と、誤分類検出手段とを備える。音響分類手段は、入力された音響信号に含まれる発話を、前記音響信号から抽出した音響的な特徴を利用して話者毎に分類し分類結果を取得する。言語特徴抽出手段は、前記発話の内容を表す文字列を取得し、この文字列を利用して前記分類結果に含まれる話者の言語的な特徴を抽出する。誤分類検出手段は、前記発話の内容を表す文字列が、前記分類結果においてこの発話が分類された話者の前記言語的な特徴に適合するか否かを判別し、適合しない場合は、この発話は前記音響分類手段によ

10

【図面の簡単な説明】

【0008】

【図1】第1の実施形態の話者分類装置を示すブロック図。

【図2】実施形態の話者分類装置のハードウェア構成を示す図。

【図3】実施形態の言語特徴ルールを示す図。

【図4】実施形態の話者分類装置のフローチャート。

【図5】実施形態の発話の分割結果を示す図。

【図6】実施形態の分類結果を示す図。

【図7】実施形態の文字列を示す図。

20

【図8】実施形態の言語的な特徴の候補を示す図。

【図9】実施形態の各発話が適合した言語特徴ルールを示す図。

【図10】実施形態の各言語特徴ルールに適合した発話数を示す図。

【図11】実施形態の各話者の言語的な特徴を示す図。

【図12】実施形態のユーザに提供する情報を示す図。

【図13】実施形態のユーザに提供する情報を示す図。

【発明を実施するための形態】

【0009】

以下、本発明の実施形態について図面を参照しながら説明する。

【0010】

30

(第1の実施形態)

第1の実施形態の話者分類装置は、会議等で収録済みの音響信号中の発話を話者毎に分類する装置である。発話を話者毎に分類することにより、議事録作成者(ユーザ)が、収録済みの音響信号のうち所望の話者の発話のみを選択的に聞くことができる。これによりユーザは、同一話者による発話内容の整合性を確認したり、会話の流れを確認したりすることが可能になり、議事録作成をスムーズに進めることができる。

【0011】

本実施形態の話者分類装置は、まず、入力された音響信号に含まれる発話を、この音響信号から抽出した音響的な特徴(例えば、MFCC)を利用して話者毎に分類し分類結果を得る。次に、各発話の音響信号を音声認識して、発話内容を表す文字列を取得する。そして、分類結果に含まれる各話者の言語的な特徴を抽出する。本実施形態の言語的な特徴としては、各話者が使用する一人称表現を表す文字列がある。この場合、ある話者(例えば、話者ID1)に分類された各発話の文字列から一人称表現(たとえば、「私は」、「僕は」など)を抽出する。そして、一人称表現毎の頻度をカウントし、頻度が所定閾値を超えるような一人称表現を話者ID1の言語的な特徴とする。

40

【0012】

次に、話者分類装置は、各発話の文字列がこの発話が分類された話者の言語的な特徴に適合するか否かを判別する。適合しない発話は、音響的な特徴を利用した分類で誤分類された発話であると判別する。例えば、発話の文字列が「僕は・・・」であり、この発話が分類された話者の一人称表現に関する言語的な特徴が「私は」である場合は、言語的な特

50

徴が適合しないため、この発話は音響的な特徴によって誤分類された発話であると判別する。

【 0 0 1 3 】

さらに、本実施形態の話者分類装置は、誤分類と判別された発話の文字列が、他の話者の言語的な特徴に適合するか否かを判別する。他の話者の言語的な特徴に適合する場合は、この発話を当該他の話者に分類する。

【 0 0 1 4 】

このように、本実施形態の話者分類装置は、言語的な特徴を用いて、音響的な特徴を利用した分類によって誤分類された発話を判別する。また、言語的な特徴を用いて、音響的な特徴によって誤分類された発話が本来分類されるべき話者を判別する。これにより、背景雑音を含んだ発話のように音響的な特徴だけでは分類が難しい発話を、言語的な特徴を用いて正しい話者に分類することができる。

10

【 0 0 1 5 】

(ブロック構成)

図 1 は、第 1 の実施形態にかかる話者分類装置を示すブロック図である。本実施形態の話者分類装置は、入力された音響信号を発話毎に分割する発話分割部 1 0 1 と、音響的な特徴を利用して分割された発話を話者毎に分類し分類結果を取得する音響分類部 1 0 2 と、音声認識を利用して各発話の内容を表す文字列を取得する文字列変換部 1 0 3 と、分類結果に含まれる各話者について発話の内容を表す文字列を取得し、この文字列を利用して各話者の言語的な特徴を抽出する言語特徴抽出部 1 0 4 と、各発話の内容を表す文字列が、分類結果においてこの発話が分類された話者の言語的な特徴に適合するか否かを判別し、適合しない場合は、この発話は前記音響分類手段によって誤った話者に分類された発話であると判別する誤分類検出部 1 0 5 と、誤った話者に分類された発話と判別された発話の文字列が、他の話者の言語的な特徴に適合するか否かを判別し、適合した場合は、この発話はこの話者に属する発話であると判別する再分類判別部 1 0 6 と、話者分類装置による分類結果をユーザに適宜提示する表示部 1 0 7 とを備える。

20

【 0 0 1 6 】

(ハードウェア構成)

本実施形態の読み記号列編集装置は、図 2 に示すような通常のコンピュータを利用したハードウェアで構成されており、装置全体を制御する C P U (Central Processing Unit) 等の制御部 2 0 1 と、各種データや各種プログラムを記憶する R O M (Read Only Memory) や R A M (Random Access Memory) 等の記憶部 2 0 2 と、各種データや各種プログラムを記憶する H D D (Hard Disk Drive) や C D (Compact Disk) ドライブ装置等の外部記憶部 2 0 3 と、ユーザの指示入力を受け付けるキーボードやマウスなどの操作部 2 0 4 と、外部装置との通信を制御する通信部 2 0 5 と、音声波形を再生して再生音を発生させるスピーカ 2 0 6 と、映像を表示するディスプレイ 2 0 7 と、これらを接続するバス 2 0 8 とを備えている。

30

【 0 0 1 7 】

このようなハードウェア構成において、制御部 2 0 1 が R O M 等の記憶部 2 0 2 や外部記憶部 2 0 3 に記憶された各種プログラムを実行することにより以下の機能が実現される。

40

【 0 0 1 8 】

(各ブロックの機能)

発話分割部 1 0 1 は、入力された音響信号を発話毎に分割する。発話毎への分割には、エネルギーやスペクトル特徴量を使用した音声区間検出技術を用いる。議事録作成対象となる音響信号は、H D D 等の外部記憶部 2 0 3 から取得したり、通信部 2 0 5 を介してネットワークで接続された他の端末 (図示なし) から取得したりすることができる。発話分割部 1 0 1 は、それぞれの発話について、「発話 I D、開始時刻、終了時刻」を取得する。ここで、発話 I D は、各発話に付与される通し番号であり、先頭から 1、2、3・・・のように付与する。開始時刻および終了時刻は、各発話の始端および終端位置が、音響信

50

号のどの時刻に対応するかを現す時間情報である。例えば、「3、00:00:18、00:00:25」という結果は、3番目（ID3）の発話が音響信号の18秒から25秒の区間に含まれることを意味する。

【0019】

音響分類部102は、音響信号から抽出した音響的な特徴を利用して、分割された発話を話者毎に分類する。音響特徴量としてはMFCCなどがある。特許文献1のように、HDD等の外部記憶部203に予め多数の話者モデルを記憶しておき、この話者モデルと各発話から抽出したMFCCとを照合した結果である類似度のパターンを利用して、各発話を話者毎に分類することができる。音響分類部102は、それぞれの発話について、「話者ID、発話ID、開始時刻、終了時刻」の組を分類結果として出力する。ここで、話者IDは、話者を識別するための識別子であり、各発話がどの話者に分類されたかを表す。発話ID、開始時刻および終了時刻には、発話分割部101の結果を用いる。

10

【0020】

文字列変換部103は、音声認識を利用して各発話の内容を表す文字列を取得する。音声認識にはMFCCなどの音響特徴量を利用することができ、MFCCとの照合に利用する音響モデル/言語モデルはHDD等の外部記憶部203に予め記憶しておく。文字列変換部103は、各発話について「発話ID、文字列」の組を出力する。ここで、文字列は各発話IDの認識結果を表している。

【0021】

言語特徴抽出部104は、音響分類部102の分類結果に含まれる各話者について、発話の内容を表す文字列を取得し、この文字列を利用して各話者の言語的な特徴を抽出する。具体的には、ある話者IDに分類された発話について、予め決められた言語特徴ルールがその発話の文字列に適合するか否かを判別する。言語特徴ルールは、過去の議事録データの形態素解析結果や一般的な辞書の内容から着目する表現（一人称表現など）を抽出することで作成できる。この他にも、人手で言語特徴ルールを作成してもよい。本実施形態において「適合」とは、文字列が一致することを意味する。

20

【0022】

本実施形態では、言語特徴ルールとして、「一人称ルール」と「文末ルール」の2種類のルールを用いる。図3に、言語特徴ルールを示す。例えば、ある話者IDに分類された発話の文字列が「私もそう思います。」であった場合、言語特徴ルールとの文字列マッチの結果、この発話の文字列は、一人称ルールID1と文末ルールID1の文字列と適合する。以上の処理を同一の話者IDに分類された他の発話についても行う。そして、各話者について同一の言語特徴ルール（一人称ルールもしくは文末ルール）に適合する発話の数をカウントし、その数が所定の閾値を超えた場合にこの言語特徴ルールを話者の言語的な特徴とする。具体的な処理は、後述のフローチャートで説明する。なお、言語特徴ルールは、HDD等の外部記憶部203に記憶しておく。

30

【0023】

誤分類検出部105は、各発話の文字列が、音響分類部102の分類結果においてこの発話が分類された話者の言語的な特徴に適合するか否かを判別する。そして、適合しない場合は、この発話は音響分類部102によって誤って分類された発話であると判別する。例えば、ある話者IDに分類された発話の文字列が「私もそう思います。」であり、この話者IDの一人称ルールに関する言語的な特徴が図3(a)の一人称ルールID2（「僕」もしくは「ぼく」）であった場合を考える。この発話の文字列は一人称ルールID2の文字列に適合しないことから、この発話は音響分類部102によって誤分類された発話と判別する。具体的な処理は、後述のフローチャートで説明する。

40

【0024】

再分類判別部106は、誤分類検出部105によって誤分類と判別された発話が、他の話者IDに属するか否かを判別する。具体的には、誤分類と判別された発話の文字列が、他の話者IDの言語的な特徴に適合するか否かを判別する。適合する他の話者IDがある場合は、この話者IDを正しい分類先の候補として出力する。例えば、誤分類と判別され

50

た発話の文字列が「私もそう思います。」である場合は、図3の一人称ルールID1および文末ルールID1を言語的な特徴として有する話者を正しい分類先候補とする。

【0025】

表示部107は、ディスプレイ207を介して話者分類装置による処理結果を適宜ユーザに提示する。

【0026】

(フローチャート)

図4は、本実施形態にかかる話者分類装置の動作を示すフローチャートである。

【0027】

ステップS401では、発話分割部101は、入力された音響信号を発話毎に分割する。図5は、発話分割部101による分割結果を表している。

10

【0028】

ステップS402では、音響分類部102は、音響的な特徴を利用して、ステップS401で分割された発話を話者毎に分類する。図6は、音響分類部102による分類結果を表している。

【0029】

ステップS403では、文字列変換部103は、分割された各発話の音響信号を音声認識して発話の内容を表す文字列を取得する。図7は、文字列変換部103によって取得された各発話の文字列を表している。

【0030】

20

ステップS404では、言語特徴抽出部104は、言語的な特徴を抽出する話者を一人選択する。ここでは、話者ID1、話者ID2・・・の順番で言語的な特徴を抽出する。

【0031】

ステップS405では、言語特徴抽出部104は、ステップS404で選択された話者に分類された発話を発話IDが小さなものから順次取得する。図6の分類結果より、話者ID1に分類された発話は、発話ID1、3、5である。ここでは、まず発話ID1の発話を取得する。

【0032】

ステップS406では、言語特徴抽出部104は、ステップS405で取得された発話の文字列が予め記憶した一人称ルールに適合するか否かを判別する。適合する場合はステップS407へ、適合しない場合はステップS408へ移行する。図7より、発話ID1の文字列は「まずは私から説明します。」である。一人称として「私」を含むことから、図3(a)の一人称ルールと文字列マッチングを行うと、一人称ルールID1が適合する。

30

【0033】

ステップS407では、言語特徴抽出部104は、選択された発話が分類された話者の言語的な特徴の候補に、ステップS406で適合した一人称ルールIDを記憶する。図8に、発話ID1、3、5の発話について記憶した一人称ルールを示す。

【0034】

ステップS408では、言語特徴抽出部104は、ステップS405で取得された発話の文字列が予め記憶した文末ルールに適合するか否かを判別する。適合する場合は、ステップS409へ、適合しない場合はステップS410へ移行する。発話ID1の場合、「ます。」が文末表現として含まれることから、図3(b)の文末ルールID1が適合する。

40

【0035】

ステップS409では、言語特徴抽出部104は、選択された発話が分類された話者の言語的な特徴の候補にステップS408で適合した文末ルールIDを記憶する。図8に、発話ID1、3、5の発話について記憶した文末ルールを示す。

【0036】

ステップS410では、言語特徴抽出部104は、ステップS404で選択された話者に分類された全ての発話について処理が終了したか否かを判別する。終了していない場合はステップS405へ移行し、他の発話について処理を行う。終了している場合はステッ

50

プ S 4 1 1 へ移行する。

【 0 0 3 7 】

ステップ S 4 1 1 では、言語特徴抽出部 1 0 4 は、全ての話者について処理が終了したか否かを判別する。終了していない場合はステップ S 4 0 4 へ移行し、他の話者について処理を行う。終了している場合はステップ S 4 1 2 へ移行する。

【 0 0 3 8 】

ステップ S 4 1 2 では、言語特徴抽出部 1 0 4 は、各話者の言語的な特徴を抽出する。具体的には、ステップ S 4 0 7 およびステップ S 4 0 9 で記憶された一人称ルールおよび文末ルールから最終的に話者の言語的な特徴とするものを決定する。図 9 に、言語的な特徴の候補として記憶された一人称ルールおよび文末ルールを示す。

10

【 0 0 3 9 】

まず、言語特徴抽出部 1 0 4 は、図 9 の言語的な特徴の候補に関する情報から図 1 0 に示すように各話者 I D について適合した一人称ルールおよび文末ルールの数をカウントする。ここで、「発話数」は各話者 I D に分類された発話総数を、「一人称ルールの適合発話数」および「文末ルールの適合発話数」は各ルールに適合した発話数を表している。

【 0 0 4 0 】

次に、各ルールに適合した発話数が所定閾値を超えるような一人称ルールもしくは文末ルールを各話者の言語的な特徴として抽出する。ここでは、同一ルール内（例えば、一人称ルール I D 1 ~ I D 3 ）における適合発話数が最大であり、かつ適合発話数を各話者 I D の発話数で正規化した値が閾値 0 . 5 以上になるようなルールを言語的な特徴として選択する。例えば、話者 I D 1 の一人称ルールの場合、一人称ルール I D 1 が最大の適合発話数（ 2 ）を示しており、かつ、話者 I D 1 の発話数で正規化した値（ 2 / 3 ）は閾値 0 . 5 以上になる。したがって、話者 I D 1 の言語的な特徴として、一人称ルール I D 1 が選択される。文末ルールについても同様な処理を適用する。以上の処理を全ての話者について実行することにより、図 1 1 に示すような話者毎の言語的な特徴を抽出できる。

20

【 0 0 4 1 】

次に、本実施形態の話者分類装置は、ステップ S 4 1 2 で抽出された話者毎の言語的な特徴を利用して、音響的な特徴によって誤分類された発話を判別する。

【 0 0 4 2 】

ステップ S 4 1 3 では、誤分類検出部 1 0 5 は、発話を 1 つ選択する。ここでは発話 I D の小さな発話から順に選択する。

30

【 0 0 4 3 】

ステップ S 4 1 4 では、誤分類検出部 1 0 5 は、選択された発話の文字列がこの発話が分類された話者 I D の言語的な特徴に適合するか否かを判別する。適合する場合はステップ S 4 1 6 へ、適合しない場合はステップ S 4 1 5 へ移行する。例えば、図 9 より発話 I D 5 は、話者 I D 1 の発話として分類されている。発話 I D 5 の文字列「僕はその意見には反対だ。」と図 1 1 の話者 I D 1 の言語的な特徴（一人称ルール I D 1 および文末ルール I D 1 ）を文字列マッチングで比較すると、一人称ルールおよび文末ルールともに適合しないことが分かる。したがって、発話 I D 5 は、音響的な特徴を使った音響分類部 1 0 2 によって誤分類された発話であると判別する。なお、本実施形態では、一人称ルールおよび文末ルールともに適合しない場合、その発話は誤って分類されたものと判別する。この他にも、一人称ルールおよび文末ルールのどちらかが適合しない場合に、誤分類された発話であると判別することもできる。

40

【 0 0 4 4 】

ステップ S 4 1 5 では、再分類判別部 1 0 6 は、ステップ S 4 1 4 で誤分類と判別された発話が本来属すべき話者を言語的な特徴を用いて判別する。具体的には、この発話の文字列が、音響分類部 1 0 2 によって分類された話者以外の他の話者の言語的な特徴に適合するか否かを判別し、適合した場合、この発話は他の話者に分類される発話であると判別する。発話 I D 5 の場合、「僕はその意見には反対だ。」の一人称は「僕は」、文末表現は「だ。」である。これらの文字列と他の話者（話者 I D 2 および話者 I D 3 ）の言語的

50

な特徴を比較すると、発話 I D 5 の文字列は話者 I D 3 の言語的な特徴（一人称ルール I D 2 および文末ルール I D 2）に適合する。したがって、発話 I D 5 は、話者 I D 3 に分類されるべき発話であると判別する。なお、本実施形態では、一人称ルールおよび文末ルールのどちらも適合した場合に再分類したが、どちらか一方が適合した場合に再分類するようにしてもよい。

【 0 0 4 5 】

このように、本実施形態の話者分類装置は、言語的な特徴を用いて、音響的な特徴を利用した分類によって誤分類された発話を判別する。また、言語的な特徴を用いて、音響的な特徴によって誤分類された発話が本来分類されるべき話者を判別する。これにより、背景雑音を含んだ発話のように音響的な特徴だけでは分類が難しい発話を、言語的な特徴を用いて正しい話者に分類することができる。

10

【 0 0 4 6 】

なお、本実施形態のステップ S 4 1 4 およびステップ S 4 1 5 では、発話の文字列と言語的な特徴の文字列マッチングを行ったが、ステップ S 4 0 7、S 4 0 9 で抽出された各発話が適合する一人称ルールおよび文末ルール（図 9）を用いて、各話者の言語的な特徴との適合性を判別してもよい。

【 0 0 4 7 】

ステップ S 4 1 6 では、全ての発話について処理が終わったか否かを判別する。終わっている場合はステップ S 4 1 7 へ、終わっていない場合はステップ S 4 1 3 へ移行する。

【 0 0 4 8 】

20

ステップ S 4 1 7 では、表示部 1 0 7 は、音響分類部 1 0 2 による分類結果および再分類判別部 1 0 6 による再分類の結果を、ディスプレイ 2 0 7 を介してユーザに提示する。図 1 2 に、ディスプレイ 2 0 7 に表示された画面の一例を示す。この図は、音響分類部 1 0 2 による分類結果を表している。「話者情報」は話者 I D を表しており、この例では「話者 I D 1」は「話者 1」と表示される。「発話内容」は音声認識によって取得された各発話の文字列を表している。図中の 1 2 0 1 はカーソルを表しており、キーボードやマウスなどの操作部 2 0 4 によりフォーカスを変更することができる。1 2 0 2 は再生ボタンであり、このボタンを押下するとカーソル 1 2 0 1 がフォーカスしている発話がスピーカ 2 0 6 から再生される。これにより、ユーザは、発話を聞いて音声認識の誤りを人手で修正することができる。また、各発話に話者情報が付与されていることから、収録済みの音響信号のうち所望の話者の発話のみを選択的に聞くことができる。これによりユーザは、同一話者による発話内容の整合性を確認したり、会話の流れを確認したりすることが可能になり、議事録作成をスムーズに進めることができる。

30

【 0 0 4 9 】

図 1 2 において、文字列に下線が引かれた発話 1 2 0 3 は、誤分類検出部 1 0 5 によって誤分類と判別された発話である。ユーザがこの発話を選択し、操作部 2 0 4 を介して所望の操作をすることにより、図 1 3 の 1 2 0 4 ように再分類判別部 1 0 6 での再分類結果をドロップダウンリストで表示させることができる。この例では、ステップ S 4 1 5 の再分類判別部 1 0 6 によって、発話 I D 5 は話者 I D 3 に再分類されていることから、「話者 3」が候補として表示される。ユーザは、再生ボタン 1 2 0 2 を押下して発話の音声聞いた後に、最終的にこの発話を話者 1 に分類するか話者 3 に分類するかを、操作部 2 0 4 を介して話者分類装置に指示することができる。

40

【 0 0 5 0 】

最後に、ステップ S 4 1 8 では、ステップ S 4 1 7 におけるユーザからの指示を取得し、最終的な話者分類を確定する。図 1 3 において、ユーザが発話 I D 5 の話者を話者 3 に変更した場合、音響分類装置は発話 I D 5 の話者は話者 I D 3 であるとして確定する。確定した話者分類の結果は、HDD等の外部記憶部 2 0 3 に記憶する。

【 0 0 5 1 】

（効果）

このように、本実施形態の話者分類装置は、言語的な特徴を用いて、音響的な特徴を

50

利用した分類によって誤分類された発話を判別する。また、言語的な特徴を用いて、音響的な特徴によって誤分類された発話が本来分類されるべき話者を判別する。これにより、背景雑音を含んだ発話のように音響的な特徴だけでは分類が難しい発話を、言語的な特徴を用いて正しい話者に分類することができる。

【 0 0 5 2 】

また、本実施形態の話者分類装置は、発話の内容を表す文字列を取得し、この文字列を利用して各話者の言語的な特徴を抽出する。これにより、各話者の言語的な特徴を予め登録しておく必要がなくなる。

【 0 0 5 3 】

(変形例 1)

本実施形態では、ステップ S 4 1 5 における分類結果をユーザに提示し、最終的な話者はユーザが選択する構成であったが、再分類判別部 1 0 6 の話者分類の結果を最終的な分類結果として確定することもできる。

【 0 0 5 4 】

また、本実施形態では、発話の文字列を音声認識で取得したが、人手で付与した文字列を発話の文字列として利用することもできる。これにより、音声認識による誤りを回避することができ、話者別の言語的な特徴をより正確に抽出できる。この場合、音響信号を話者別に分類する話者インデキシング等にも本実施形態を適用することができる。

【 0 0 5 5 】

なお、以上説明した本実施形態における一部機能もしくは全ての機能は、ソフトウェア処理により実現可能である。

【 0 0 5 6 】

本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

【 符号の説明 】

【 0 0 5 7 】

- 1 0 1 発話分割部
- 1 0 2 音響分類部
- 1 0 3 文字列変換部
- 1 0 4 言語特徴抽出部
- 1 0 5 誤分類検出部
- 1 0 6 再分類判別部
- 1 0 7 表示部
- 2 0 1 制御部
- 2 0 2 記憶部
- 2 0 3 外部記憶部
- 2 0 4 操作部
- 2 0 5 通信部
- 2 0 6 スピーカ
- 2 0 7 ディスプレイ
- 2 0 8 バス
- 1 2 0 1 カーソル
- 1 2 0 2 再生ボタン
- 1 2 0 3 発話内容
- 1 2 0 4 ドロップダウンリスト

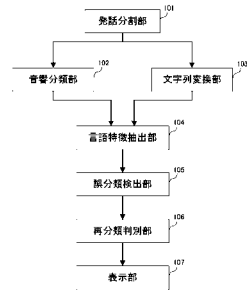
10

20

30

40

【 図 1 】



【 図 3 】

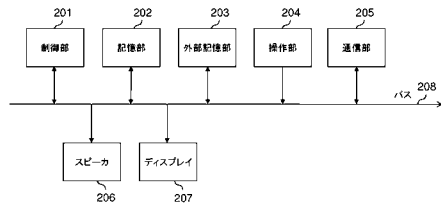
一人称ルールID	一人称ルール
1	私、わたし
2	僕、ぼく
3	俺、おれ

(a)

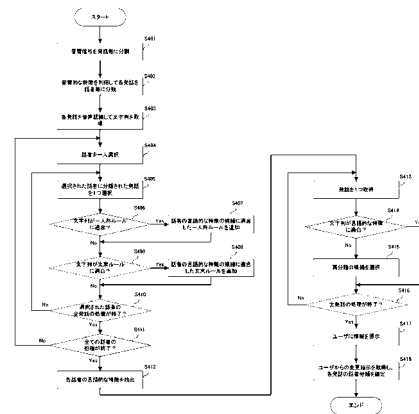
文末ルールID	文末ルール
1	です。、ます。
2	だ。、である。

(b)

【 図 2 】



【 図 4 】



【 図 5 】

発話ID	開始時刻(時:分:秒)	終了時刻(時:分:秒)
1	00:00:00	00:00:02
2	00:00:03	00:00:06
3	00:00:10	00:00:15
4	00:00:18	00:00:25
5	00:00:30	00:00:35
6	00:00:36	00:00:40

【 図 7 】

発話ID	文字列
1	まずは私から説明します。
2	よろしく願います。
3	私は賛成です。
4	私もそう思います。
5	僕はその意見には反対だ。
6	僕が今まで考えてきたことはこうだ。

【 図 6 】

話者ID	発話ID	開始時刻(時:分:秒)	終了時刻(時:分:秒)
1	1	00:00:00	00:00:02
2	2	00:00:03	00:00:06
1	3	00:00:10	00:00:15
2	4	00:00:18	00:00:25
1	5	00:00:30	00:00:35
3	6	00:00:36	00:00:40

【圖 8】

話者ID	発話ID	文字列	言語的な特徴の候補	
			適合した一人称ルールID	適合した文末表現ルールID
1	1	まずは私から説明します。	1	1
1	3	私は賛成です。	1	1
1	5	僕はその意見には反対だ。	2	2

【 図 9 】

話者ID	発話ID	文字列	言語的な特徴の候補	
			適合した一人称ルールID	適合した文末表現ルールID
1	1	まずは私から説明します。	1	1
2	2	よろしくお願いします。	non	1
1	3	私は賛成です。	1	1
2	4	私もそう思います。	1	1
1	5	僕はその意見には反対だ。	2	2
3	6	僕が今まで考えてきたことはこうだ。	2	2

【 図 1 1 】

話者ID	言語的な特徴	
	一人称ルールID	文末表現ルールID
1	1	1
2	1	1
3	2	2

【 図 1 0 】

話者ID	発話数	一人称ルールの適合発話数			文末表現ルールの適合発話数		
		ID1	ID2	ID3	ID1	ID2	ID3
1	3	2	1	0	2	1	0
2	2	1	0	0	2	0	0
3	1	0	1	0	0	1	0

【 図 1 2 】

話者情報		発話内容	
話者1	まずは私から説明します。		1201
話者2	よろしくお願いします。		
話者1	私は賛成です。		
話者2	私もそう思います。		1203
話者1	僕はその意見には反対だ。		
話者3	僕が今まで考えてきたことはこうだ。		
			1202

【 図 1 3 】

話者情報		発話内容	
話者1	まずは私から説明します。		
話者2	よろしくお願いします。		
話者1	私は賛成です。		1201
話者2	私もそう思います。		1203
話者1 ▽	僕はその意見には反対だ。		
話者3	僕が今まで考えてきたことはこうだ。		
			1202

フロントページの続き

- (72)発明者 池田 朋男
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 永尾 学
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 西山 修
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 鈴木 博和
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 上野 晃嗣
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 下郡 信宏
東京都港区芝浦一丁目1番1号 株式会社東芝内