

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2009年8月27日(27.08.2009)

(10) 国際公開番号
WO 2009/104332 A1

- (51) 国際特許分類:
G10L 15/04 (2006.01) G10L 15/28 (2006.01)
G10L 15/00 (2006.01) G10L 17/00 (2006.01)
G10L 15/22 (2006.01)
- (21) 国際出願番号: PCT/JP2008/072851
- (22) 国際出願日: 2008年12月16日(16.12.2008)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2008-037200 2008年2月19日(19.02.2008) JP
- (71) 出願人(米国を除く全ての指定国について): 日本電気株式会社(NEC Corporation) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人(米国についてのみ): 長友 健太郎(NAGATOMO, Kentaro) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).
- (74) 代理人: 家入 健(IEIRI, Takeshi); 〒2210835 神奈川県横浜市神奈川区鶴屋町三丁目3番8ア

サヒビルディング 10階 響国際特許事務所 Kanagawa (JP).

- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

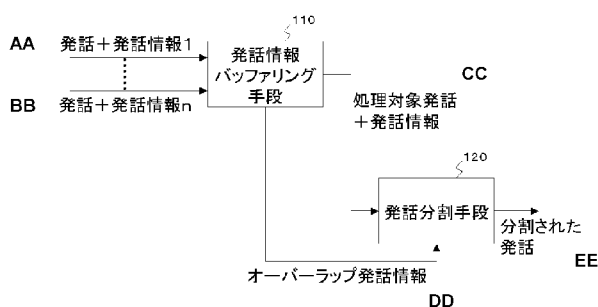
添付公開書類:

- 国際調査報告 (条約第 21 条(3))

(54) Title: SPEECH SEGMENTATION SYSTEM, SPEECH SEGMENTATION METHOD, AND SPEECH SEGMENTATION PROGRAM

(54) 発明の名称: 発話分割システム、発話分割方法および発話分割プログラム

[図3]



AA SPEECH + SPEECH INFORMATION 1
 BB SPEECH + SPEECH INFORMATION n
 110 SPEECH INFORMATION BUFFERING MEANS
 CC PROCESSING TARGET SPEECH + SPEECH INFORMATION
 120 SPEECH SEGMENTATION MEANS
 DD OVERLAPPING SPEECH INFORMATION
 EE SEGMENTED SPEECH

(57) Abstract: In a dialogue sound in which temporally overlapped speeches may occur, in order to segment speeches at appropriate positions to facilitate better understanding of a relationship between the speeches, a speech segmentation system comprises a speech segmentation means (120) for receiving speeches by two or more speakers and further segmenting each of the speeches. If another speech is made so as to overlap a certain speech, it is considered that a time point when the overlapping speech is spoken becomes a semantic segmentation. Therefore, the speech segmentation means (120), when determining a segmentation point of a certain speech, uses information on another speech that overlaps.

(57) 要約: 時間的にオーバーラップする発話が起こり得るような対話音声において、各発話の関連性がより把握しやすいように、発話を適切な位置で分割するため、二人以上の話者による発話を入力とし、それぞれの発話をさらに分割する発話分割手段 120 を備える。ある発話に対して他の発話がオーバーラップするように発話されたなら、そのオーバーラップする発話が発せられた時点が当該発話における一つの意味的な切れ目になっていると考えられる。この

ため、発話分割手段 120 は、ある発話の分割点の決定に際し、オーバーラップする他の発話の情報を
用いる。

WO 2009/104332 A1

明 細 書

発話分割システム、発話分割方法および発話分割プログラム

技術分野

[0001] 本発明は、発話分割システム、発話分割方法および発話分割用プログラムに関し、特に、二人以上の話者による対話音声を入力とし、ある話者が発した発話について、他の話者によってその発話と時間的にオーバーラップするような発話が為された場合であっても、対話の流れが読み取りやすいように当該発話を適切に分割することを特徴とする発話分割システム、発話分割方法および発話分割プログラムに関する。

背景技術

[0002] 音声認識技術は、これまで主に話者が一名である場合を想定して研究、開発が進められてきた。

[0003] 近年、音声認識技術を対話音声へ適用した対話音声認識システムがいくつか登場してきている。これらは、基本的には一名の話者に対して行っていた処理を、対話に参加する個々の話者ごとに施すだけのものが多い。

[0004] 今日の一般的な音声認識技術において、処理の最小単位は「発話」である。音声認識システムは、入力された音声信号を発話ごとに分割し、各発話について音声認識結果を(主にテキストとして)出力し、その結果を順に画面や記憶装置等へ出力する。

[0005] ここで発話とは、ある程度の長さを持った音声の塊を指す。直感的には、一発話はほぼ一つの文(またはそれに準ずる発話内容の意味的な単位)に相当する。

[0006] 一般的な音声認識技術では、入力音声信号のうち、長い無音区間を含まない音声区間を発話として切り出すことで発話単位を切り出す。これは、比較的短い文であれば、ほぼ一つの発話に収める(その文の内容を音声として発声し切る)ことができるといふ知見に基づく。長い文の場合は、複数の発話に分かれることも多いが、これはそのような内容の発話を行う場合には息継ぎをすることがあるためである。

[0007] 発話を検出するアルゴリズムを記載した文献として、非特許文献1がある。非特許文献1では、ある一定時間内において、音声の振幅レベルが指定値以上の振幅の零交

差数が一定数以上になれば音声区間開始(発話開始)、同数が一定値以下になれば発声終了(発話終了)とみなす。

[0008] 一方で、通常、入力音声を発話単位に分割する処理は発話の内容を参照せずに施されるため、複数の文に相当する音声区間が一塊の発話として切り出されることもしばしば起こり得る。また、自然発話の場合では、文法的な意味での「文」という単位自体がそもそもあまり意識されないので、自ずと発話と文の対応関係も曖昧になされることが多い。

[0009] このように、音声認識技術の処理単位である「発話」は、「文」(またはそれに準ずる発話内容の意味的な単位)を基本とするが、「文」と等価になることは保証されていない、という性質を持つ。

[0010] このような従来の音声認識システムを対話音声に適用する場合、各話者の音声それぞれについて発話への分割が行われ、その各々について音声認識処理を施すことになる。

[0011] 従来の対話音声認識システムの一例が、特許文献1に記載されている。この従来の対話音声認識システムは、複数話者が発したそれぞれの音声から発話を切り出す発話検出手段と、切り出された発話を切り出された順に処理する音声認識手段と、その結果を順に画面に表示する出力手段から構成される。

[0012] このような構成を有する従来の対話音声認識システムは、つぎのように動作する。

[0013] すなわち、発話検出手段は、複数の話者が発したそれぞれの音声信号から発話を検出する。検出された発話は、検出された順に(つまり発話開始時刻の順に)音声認識手段に送られ、音声信号からテキスト等の認識結果に変換される。最終的に認識結果は出力手段に送られ、音声認識手段が処理した順に(つまり発話検出順に)画面等に表示される。

[0014] この方式は、会議における音声対話のように、ある時点における主たる発話者同士の発話内容が互いに独立している場合には問題なく動作する。

[0015] 例えば、統制の取れた会議の場合には、ほとんどすべての時刻においてたかだか一名のみが発話を行う。参加者は、アイコンタクトや挙手等で自ら発話権を主張するか、あるいは他の参加者の催促によって発話権の取得を促され、最終的に議長の許

可をもって発話権の移譲を受けたのち、初めて発話が許される。例外は発話権の取得を主張する為の発話であるが、これも通常は現に発話権を持つ話者の発話を妨害しないことが求められる。このようなスタイルの会議としては、国会や地方議会などのフォーマルな会議が挙げられる。

[0016] このような対話では、ある話者による発話と別の話者による発話とは必ず時間的にオーバーラップしない。従って、認識結果テキストを、発話の現れた順にただ出力するだけの特許文献1のようなシステムであっても、最終的に出力された対話全体の認識結果は、ある程度理解しやすいものとなる。

[0017] しかしながら、よりインフォーマルな会議や、あるいは通常の日常会話、電話対話などでは、複数話者による発話のオーバーラップが頻繁に生じる。

[0018] 特に電話対話(電話越しの音声対話)ではこのオーバーラップが顕著に現れる。これは、アイコンタクト等の非音声コミュニケーション手段を用いることが出来ないという理由による。

[0019] 非音声コミュニケーション手段が利用できないため、電話対話での発話権のやり取りは音声発話をもって為される。この発話権要求発話を発するタイミングは、概ね相手の発話の切れ目を狙って行われると期待できるが、実際には発話の切れ目を上手く推察できない場合が頻発する。また、互いに相手の発話の切れ目であると判断して同時に発話を開始するケースも良く見受けられる。

[0020] また、発話の中にはあいづちや確認など、必ずしも発話権を伴わないものも存在する。これらは対面会話においてはアイコンタクトやうなづきなどで代替可能なものもあるが、電話対話では音声発話で行うしかなく、発話のオーバーラップが増える一因となっている。

[0021] フォーマルな会議における対話と、それ以外の対話でのもう一つの違いは、インフォーマルな対話では、発話権を要求する発話(例えば「議長」「よろしいですか」「すみません」)などが省略されるケースが見られるという点である。これは、非音声コミュニケーション手段によってある程度代替が可能である(例えば挙手した後、誰も反対でないと目視で確認できる場合など)というだけでなく、発話のオーバーラップが生じて改めて発話しなおすというプロトコルを用いるよりも、発話権要求と発話内容の伝達

を同時に行った方が、長い目で見て対話全体の時間削減に繋がるという個々人の経験則に基づくものと考えられる。

[0022] このように、インフォーマルな対話や電話越しの音声対話などでは、ある話者による発話と別の話者による発話とが時間的にオーバーラップすることがあり得る。このような場合、特許文献1のように、発話の現れた順に認識結果テキストを出力するだけでは、最終的に出力された対話全体の認識結果は、非常に見づらいものになってしまう。

[0023] 以下、図1および図2を用いて、この理由を説明する。

[0024] 図1は、2名の話者が互いにあいづちや確認を交えながら交互に発話している様子を人手で書き起こしたものである。表は、対話の流れが明確になるようその内容を考慮しながら、各話者の発話を適切な時点で分割して書き起こしたものである。このように示すことで、話者AおよびBがそれぞれ互いのどの発話内容に対してあいづちや同意を示しているかが良く分かる。一方、その下に示した時系列図を見ると、このように理想的に発話を切り出した場合でも、一部の発話がオーバーラップしている様子が分かる。例えば発話番号1-4,1-5,1-6はそれぞれオーバーラップして発話されている。

[0025] この対話音声を従来法に基づく音声認識システムに与えた場合の一例を図2に示す。図2では、従来の音声認識手段を用い、発話内容と無関係に発話の切り出し処理を行っている。その結果、図1では別個の「発話」であったものが、一つの「発話」にまとめられた例が散見される。例えば図2の発話2-1は、図1の発話1-1,1-3,1-5,1-7を合わせたものである。何故、こうなるかと言えば、前述したように一般的な音声認識技術では、入力音声信号のうち、長い無音区間を含まない音声区間を一つの発話として切り出すためである。同様に、話者Aの発話1-9,1-11はそれらの間の無音区間の長さが短いため一つの発話として切り出され、話者Bの発話1-4,1-6はそれら間の無音区間の長さが短いため一つの発話として切り出され、話者Bの発話1-8,1-10はそれらの間の無音区間の長さが短いため一つの発話として切り出される。

[0026] ここで、時刻20:30:17.0に話者Bが発した「はい」という発話が、話者Aのどの発話内容に対するものであるかを知りたいとする。これは図1のような理想的な出力であれば

自明である(直近の話者Aの発話1-1)が、図2では発話2-1のどの内容であるかは自明ではない。このケースでは、この「はい」は、話者Bが「先ほどお伺いした件」を了解したことを示すものかも知れないし、「連絡いただけてない」という状況を理解したことを示すものであるかも知れないし、あるいは話者Aの「はい」に対する単なる拍子合わせのためだけのものである可能性すら考えられる。

[0027] また、時刻20:30:22.8から話者Bの問い合わせ「それは、1時ごろの」に対する話者Aの回答は、図1であれば発話1-11の「いえ、～」で始まる否定文であることが一目で分かるが、図2では発話2-5のように肯定語「ええ」と否定語「いえ」が連続しており、不明瞭である。

[0028] なお、図2では発話の順序関係に注目するために音声認識そのものは正しく行われたとして記載しているが、実際の認識結果には誤りが含まれるため、さらに対話の流れが読み取りにくいことを付記しておく。

[0029] 以上の例が示すように、インフォーマルな対話や、電話対話などの音声発話のみを用いる対話においては、複数話者による発話のオーバーラップが頻繁に生じ、これを従来手法のように長い無音区間を含まない音声区間を一つの発話として切り出す単純な音声認識手段によって処理すると、対話の流れを把握しづらくなり、読みづらい出力となってしまふ。この例では対話音声認識結果の表示のみを問題として取り上げたが、同様の問題は、例えば対話内容を解析するような例(上述の「それは、1時ごろの」に対する返答の例)などにも影響する。

[0030] 特許文献1:特許第3859612号

非特許文献1:”Juliusにおける入力の無音区間・休止の扱い”[平成20年2月2日検索]インターネット<URL:<http://julius.sourceforge.jp/index.php?q=doc/iwsp.html>>

非特許文献2:瀬戸口、高梨、河原,多数のセンサを用いたポスター会話の収録とその分析,情報処理学会研究報告SIG SLP 2007-SLP-67

発明の開示

発明が解決しようとする課題

[0031] 第1の問題点は、二人以上複数の話者が参加する音声対話において、二人以上

の話者が発した発話が時間的にオーバーラップする場合、従来法のように長い無音区間を含まない音声区間を一つの発話として切り出す素朴な方法に基づく対話音声認識システムの出力は、対話における各発話の関連性が把握しづらくなり、ひいては、音声対話全体の認識結果の読解や自然言語处理的分析が困難に成る、という点である。

[0032] [発明の目的]

本発明の目的は、各発話の関連性がより把握しやすい形態に容易に出力を整形できるように、発話を適切な位置で分割することにある。

課題を解決するための手段

[0033] 本発明の発話分割システムは、二人以上の話者による発話を入力とし、それぞれの発話をさらに分割する発話分割手段を備える。前記発話分割手段は、ある発話の分割点の決定に際し、オーバーラップする他の発話の情報を用いる。

[0034] このような構成を採用し、ある発話の分割を試みる際、その発話と時間的にオーバーラップする他の発話の情報を用いて、当該発話の分割位置を決定し、分割することによって、本発明の目的を達成することができる。

発明の効果

[0035] 第1の効果は、時間的にオーバーラップする発話が起こり得るような対話音声においても、それぞれの発話を、音声対話中の各発話の関連性がより把握しやすい適切な位置でさらに分割することができることである。

[0036] その理由は、対話に参加する各話者は、原則として他者の発話をできるだけ遮らないように、遮るとしてもできるだけ対話の流れを混乱させないように発話する傾向があるため、ある発話に対して他の発話がオーバーラップするように発話されたなら、そのオーバーラップする発話が発せられた時点が当該発話における一つの意味的な切れ目になっていると考えられるためである。

図面の簡単な説明

[0037] [図1]本発明の効果を説明するために示した、二人の話者による対話の例を示す図である。

[図2]本発明の効果を説明するために示した、図1の対話を従来法を用いて音声認

識した認識結果の出力例を示す図である。

[図3]本発明の第1の実施の形態の構成を示すブロック図である。

[図4A]本発明の第2の実施の形態の構成を示すブロック図である。

[図4B]本発明の第2の実施の形態における発話分割手段の構成を示すブロック図である。

[図5]本発明の第3の実施の形態の構成を示すブロック図である。

[図6]本発明の第4の実施の形態の構成を示すブロック図である。

[図7]本発明の実施例の構成を示すブロック図である。

符号の説明

- [0038] 110…発話情報バッファリング手段
120、120A…発話分割手段
121…発話ブロック化手段
122…分割点尤度保持・更新手段
123…分割点尤度特徴量抽出手段
124…分割点決定手段
210…発話検出手段
220…発話ブロック化手段
230…発話バッファリング手段
240…分割点尤度特徴量抽出手段
250…分割点尤度更新手段
260…分割点決定手段
310…発話検出手段
320…発話ブロック化手段
330…音声認識手段
340…発話バッファリング手段
350…分割点尤度特徴量抽出手段
360…分割点尤度更新手段
370…分割点決定手段

- 410…音声入力デバイス
- 420…発話検出モジュール
- 430…音声認識モジュール
- 440…発話情報バッファ
- 450…単語アレイバッファ
- 460…分割点尤度更新モジュール
- 470…認識結果分割モジュール
- 480…認識結果整列モジュール
- 490…表示装置

発明を実施するための最良の形態

- [0039] 次に、発明を実施するための最良の形態について図面を参照して詳細に説明する。
- [0040] 図3を参照すると、本発明の第1の実施の形態は、発話情報バッファリング手段110と、発話分割手段120とから構成されている。これらの手段はそれぞれ概略次のように動作する。
- [0041] 発話情報バッファリング手段110は、発話およびその発話に関する情報(発話情報)を受け取り、一時的に保持する。ここで発話情報とは、例えば発話の開始・終了時刻などを含む。
- [0042] 発話分割手段120は、発話情報バッファリング手段110から未処理の発話を一つ取り出し、その発話に対して分割処理を試みる。このとき、その発話の発話情報を参照し、この発話とオーバーラップする発話の発話情報を発話情報バッファリング手段110から取り出して利用する。
- [0043] 次に、図3を参照しつつ本実施の形態の全体の動作について詳しく説明する。
- [0044] まず、処理対象となる発話データが、発話情報とあわせて発話情報バッファリング手段110に入力される。
- [0045] 次に、発話分割手段120は、発話情報バッファリング手段110から、まだ発話分割処理が行われておらず、かつ発話開始時刻がもっとも早い発話の一つを選び出し、処理対象発話とする。

- [0046] さらに、発話分割手段120は、発話情報バッファリング手段110に蓄積された各発話の発話情報を参照し、処理対象発話と時間的にオーバーラップする発話を見つけ出す。
- [0047] ここで、処理対象発話とオーバーラップする発話が一つもなければ、発話分割手段120は処理対象発話の分割に失敗したとして処理対象発話をそのまま出力し、一方、処理対象発話とオーバーラップする発話の一つ以上あった場合、それらオーバーラップ発話の発話情報を用いて、処理対象発話を分割する。具体的には、オーバーラップする各発話の開始時刻で、処理対象発話を分割する。
- [0048] すべてのオーバーラップする発話を評価した時点で、処理対象発話の分割が完了する。発話分割手段120は、処理対象発話の分割結果を時刻順に出力する。
- [0049] 以上の処理を、発話情報バッファリング手段110に未処理の発話なくなるまで繰り返す。
- [0050] 次に、本実施の形態の効果について説明する。
- [0051] 本実施の形態では、分割処理の対象となる発話を発話情報バッファリング手段110に蓄積することで、ある発話に対して時間的にオーバーラップする他の発話を確実に発見することができる。
- [0052] なお、本実施の形態は、既に与えられた発話を、バッチ的に処理して発話分割を試みるのに適した形態である。
- [0053] すなわち、予め何らかの方法で発話に分割されていた音声が発話情報バッファリング手段110に蓄積されており、これらを一度に与えることで、それら一つ一つを適切に再分割するような用途に向く。
- [0054] 本実施の形態を、非バッチ的なオンライン処理で用いる場合は、適当な遅延時間を定め、その遅延時間ごとに発話情報バッファリング手段110に発話をバッファリングする。ただしこの場合、ある発話に対してオーバーラップする発話が実際には存在したとしても、その発話の検出確定が遅延時間内に行われなかった場合、そのオーバーラップ発話を考慮せずに分割することとなり、若干の分割精度の低下が起り得る。
- [0055] 次に、本発明の第2の実施の形態について図面を参照して詳細に説明する。
- [0056] 図4Aを参照すると、本発明の第2の実施の形態は、図3に示した第1の実施の形

態と比較して、発話分割手段120の代わりに発話分割手段120Aを備えている点で相違する。

- [0057] 図4Bを参照すると、本発明の第2の実施の形態における発話分割手段120Aは、発話ブロック化手段121と、分割点尤度保持・更新手段122と、分割点尤度特徴量抽出手段123と、分割点決定手段124を含む。これらの手段はそれぞれ概略つぎのように動作する。
- [0058] 発話ブロック化手段121は、入力された分割処理対象発話を、後続の処理に適したブロックに分割する。
- [0059] 分割点尤度保持・更新手段122は、処理対象発話の各ブロックについて、各ブロックが、処理対象発話の分割点となりうる尤度(分割点尤度)を保持し、また、更新する。
- [0060] 分割点尤度特徴量抽出手段123は、処理対象発話と時間的にオーバーラップする発話を解析し、処理対象発話の各ブロックの分割点尤度を更新するために用いられる分割点尤度特徴量を抽出する。
- [0061] 分割点決定手段124は、分割点尤度保持・更新手段122に保持される処理対象発話の各ブロックの分割点尤度を評価し、最終的に採用する分割点を決定し、処理対象発話をその分割点で分割して出力する。
- [0062] 次に、図4Aおよび図4Bを参照して本実施の形態の全体の動作について詳細に説明する。
- [0063] まず、第1の実施の形態と同様に、処理対象となる発話データが、発話情報とあわせて発話情報バッファリング手段110に入力される。次に、発話分割手段120Aは、発話情報バッファリング手段110から、まだ発話分割処理が行われておらず、かつ発話開始時刻がもっとも早い発話を一つ選び出し、処理対象発話とする。そして、まず、新たな処理対象発話が、その発話の発話情報とともに、発話ブロック化手段121に入力される。
- [0064] 発話ブロック化手段121は、入力された処理対象発話を定められた方法で分割し、ブロック化する。これは、例えば10ミリ秒単位で等分する、などの方法で行われる。
- [0065] 次に、分割点尤度保持・更新手段122は発話ブロック化手段121が出力した処理

対象発話の各ブロックについて、分割点尤度はゼロである(分割点ではない)として初期化する。

- [0066] ここで、発話情報バッファリング手段110から、処理対象発話と時間的にオーバーラップする発話の発話情報が、分割点尤度特徴量抽出手段123に逐次入力される。すると、分割点尤度特徴量抽出手段123は、定められたアルゴリズムに従ってオーバーラップ発話を解析し、分割点尤度特徴量を得る。具体的には、分割点尤度特徴量として、オーバーラップ発話の開始時刻、終了時刻、発話継続時間長、オーバーラップ時間などが抽出される。
- [0067] こうして得られた分割点尤度特徴量は、順次、分割点尤度保持・更新手段122に入力される。分割点尤度保持・更新手段122はこれらの特徴量を用いて処理対象発話の各ブロックに対する分割点尤度を、定められたアルゴリズムに従って更新する。
- [0068] 例えば、オーバーラップする発話の開始時刻に相当するブロックやその前後のブロックについては分割点尤度が高くなるよう更新する。
- [0069] その一方で、オーバーラップ発話の継続時間長が非常に長い場合や、処理対象発話とオーバーラップ発話のオーバーラップ時間が大きい場合には、開始時刻相当ブロックへの分割点尤度の更新は控えめに行う。これは、オーバーラップ発話の話者が、処理対象発話を見逃して発話している可能性が高く、処理対象発話の分割点を求める基準としては用いない方が適切であるという経験則を反映したものである。
- [0070] また別の例としては、オーバーラップ発話の継続時間長が非常に短い場合、そのオーバーラップ発話はノイズなどの誤検出である可能性があるため、分割点尤度の更新は控えめに行うようにする。
- [0071] またさらに別の例として、もし発話情報としてその発話の非音声尤度が与えられているのであれば、オーバーラップ発話の開始時刻の直近で非音声尤度の高いブロックに対して特に高い分割点尤度を当てるようにしても良い。ここで、非音声尤度とは、ある音声の微少な1区間が雑音モデルにどの程度マッチしているかを示す確度を意味する。このような非音声尤度は、一般の音声認識処理において微少な音声区間毎に求められている。
- [0072] すべてのオーバーラップする発話を評価した時点で、分割点尤度保持・更新手段1

22には、処理対象発話の各ブロックに対する分割点尤度が求まっている。分割点決定手段124はこの情報を元に、定められたアルゴリズムを用いて最終的な分割点を決定する。例えば、予め与えられた閾値を超える分割点尤度を持つブロックを分割点とする。一つの発話に対する最大分割数に上限を設けてもよい。分割点決定手段124は、処理対象発話の分割結果を時間順に出力する。開始時刻が同じ場合には、終了時刻の早い順に出力する。

[0073] なお、処理対象発話に対するオーバーラップ発話が一つも見つからなかった場合、分割点尤度保持・更新手段122に保持される各ブロックの分割点尤度はゼロのままである。この場合、分割点決定手段124は分割点が一つも見つからなかったとして、もとの処理対象発話をそのまま出力する。

[0074] ここで、分割点決定手段124は、分割された処理対象発話を発話情報バッファリング手段110に改めて入力しても良い。この場合、これらの分割された発話は未処理扱いとし、一方、分割前の処理対象発話は発話情報バッファリング手段110から削除する。

[0075] 発話分割手段120Aは、以上の処理を、発話情報バッファリング手段110に未処理の発話がなくなるまで繰り返す。

[0076] 次に、本実施の形態の効果について説明する。

[0077] 本実施の形態では、処理対象発話の分割点を求めるに際して、処理対象発話の各部における分割点尤度を判定することによって分割点を決定するように構成されている。このため、オーバーラップ発話のさまざまな情報を複雑に混交して、より精度の高い分割点推定を行うことができる。

[0078] また、本実施の形態では、さらに、発話分割手段120Aによって一度分割された発話を再び発話情報バッファリング手段110に登録するため、さらに精度の高い発話分割が可能となる。なぜなら、発話情報バッファリング手段110に登録されている各発話は潜在的にさらに分割可能であり、発話を分割すると、分割点尤度特徴量としてのオーバーラップ発話の終了時刻、発話継続時間長、オーバーラップ時間などが増えるためである。

[0079] なお、本実施の形態も第1の実施の形態と同じく、既に与えられた発話を、バッチ的

に処理して発話分割を試みるのに適した形態である。本実施の形態を、非バッチ的なオンライン処理で用いる場合は、適当な遅延時間を定め、その遅延時間ごとに発話情報バッファリング手段110に発話をバッファリングする。こうすることで、再分割された発話をその遅延時間だけの遅れで得ることが出来る。ただしこの場合、ある発話に対してオーバーラップする発話が実際には存在したとしても、その発話の検出確定が遅延時間内に行われなかった場合、そのオーバーラップ発話を考慮せずに分割することとなり、若干の分割精度の低下が起こり得る。

[0080] 次に、本発明の第3の実施の形態について図面を参照して詳細に説明する。

[0081] 図5を参照すると、本発明の第3の実施の形態は、発話検出手段210と、発話ブロック化手段220と、発話バッファリング手段230と、分割点尤度特徴量抽出手段240と、分割点尤度更新手段250と、分割点決定手段260とから構成されている。これらの手段はそれぞれ概略次のように動作する。

[0082] 発話検出手段210は、複数の話者によって同時に発された一つ以上複数の音声を入力とし、それぞれについて定められたアルゴリズムを用いて音声中の発話を検出し、その開始時刻および終了時刻を少なくとも含む、発話情報を出力する。また発話区間の音声信号も同時に出力される。

[0083] 発話を検出するアルゴリズムは既知の様々な手法を用いることが出来る。例えば前述した非特許文献1に記載された手法を用いることができる。

[0084] 発話ブロック化手段220は、入力された分割処理対象発話を、後続の処理に適したブロックに分割する。

[0085] ただし、多くの場合、発話検出手段210は入力音声を何らかの方法でブロック化するので、このブロックをそのまま流用してもよく、その場合、発話ブロック化手段220は省略できる。発話検出手段210で見られるブロック化としては、例えばサンプリング周期ごとの時間区間への分割(これはつまり音声信号の量子化単位そのものである)や、音声信号をフーリエ変換等の信号処理に掛ける単位(一般にフレームと呼ばれる)などが挙げられる。

[0086] 発話バッファリング手段230は、各発話の発話情報と、発話そのものを保持する。また、第2の実施の形態において分割点尤度保持・更新手段122が保持していた、発

話の各ブロックに対する分割点尤度情報も同時に保持する。

- [0087] 分割点尤度特徴量抽出手段240は、第2の実施の形態における分割点尤度特徴量抽出手段123と基本的に同等の動作をする。
- [0088] 分割点尤度更新手段250は、分割点尤度特徴量抽出手段240によって得られた特徴量を用いて、発話バッファリング手段230が保持する各発話の各ブロックに対する分割点尤度を更新する。
- [0089] 分割点決定手段260は、発話バッファリング手段230に保持される一つまたは複数の処理対象発話の各ブロックの分割点尤度を評価し、最終的に採用する分割点を決定し、各処理対象発話をその分割点で分割して出力する。
- [0090] 次に、図5を参照して本実施の形態の全体の動作について詳細に説明する。
- [0091] まず、二つ以上複数の音声信号が、発話検出手段210に入力される。発話検出手段210は、この入力音声順次処理し、各時刻において、各音声で発話が始まっているか否かを検証する。また、既に発話の開始を検出した音声については、その発話が終了しているか否かを検証する。いずれかの音声について、発話検出手段210が発話の開始を検出すると、その時刻が発話バッファリング手段230に送られる。また同様に、発話の終了時刻も発話バッファリング手段230に送られる。
- [0092] 発話検出手段210で処理された音声信号は、発話ブロック化手段220を通じてブロック化され、発話バッファリング手段230に送られる。既に述べたように、発話検出手段210が何らかのブロック化を行う場合、発話ブロック化手段220を省略しても構わない。
- [0093] 発話バッファリング手段230は、時間軸方向および話者数について可変な、ブロック化された発話の二次元配列で構成される。
- [0094] 新たなブロックが発話バッファリング手段230に入力されると、二次元配列の相当する時刻・話者の位置が確保される。もし当該ブロックの時刻が発話開始時刻以降かつ発話終了時刻以前であるなら、そのブロックに対する分割点尤度の初期値であるゼロが、二次元配列のその位置にセットされる。
- [0095] 分割点尤度特徴量抽出手段240は、適当なタイミングで呼び出され、発話バッファリング手段230中の未処理のブロックについて、定められたアルゴリズムに従って解

析し、分割点尤度特徴量を得る。さらに分割点尤度更新手段250がこの特徴量を元に各ブロックの分割点尤度を定められたアルゴリズムに従って更新する。以下に具体的な例の詳細を示す。

- [0096] 第1の例は、オーバーラップする発話の開始時刻を特徴量として抽出するケースである。この場合、分割点尤度特徴量抽出手段240は、まず、発話バッファリング手段230が保持する各発話の開始時刻のうち、未だ処理していないものを時刻の早いものから一つ取り出す。次に、その時刻に相当する各発話のブロックについて、分割点尤度更新手段250が、分割点尤度を高くなるように更新する。
- [0097] 第2の例は、オーバーラップする発話の開始時刻の前後の分割点尤度も更新するケースである。この場合、分割点尤度特徴量抽出手段240が起動するタイミングは、未処理のある発話開始時刻に対して、尤度更新を反映する範囲の終端に相当するブロックが、発話バッファリング手段230に登録された後になる。分割点尤度更新手段250は、処理対象となる開始時刻に相当する各ブロックと、その前後のブロックに対して、尤度を高めるよう更新する。
- [0098] 第3の例は、オーバーラップする発話の継続時間に応じて、尤度更新の効果を制御するケースである。この場合、予めタイムアウト時間を定め、分割点尤度特徴量抽出手段240が起動するタイミングは、未処理のある発話開始時刻に対してこのタイムアウト時間が経過した時刻に相当するブロックが、発話バッファリング手段230に登録された後になる。分割点尤度特徴量抽出手段240は、発話バッファリング手段230に保持された情報から、処理対象となる開始時刻から始まる発話の継続時間長を求め、分割点尤度更新手段250は、当該発話開始時刻に相当する各発話のブロックについて分割点尤度を高くなるように更新するが、この際の尤度更新は、分割点尤度特徴量抽出手段240によって得られたオーバーラップ発話の継続時間長に応じて調整する。具体的には第2の実施の形態と同様に、オーバーラップ発話の継続時間長が非常に長い場合には、開始時刻相当ブロックへの分割点尤度の更新は控えめに行い、また、オーバーラップ発話の継続時間長が非常に短い場合には、そのオーバーラップ発話はノイズなどの誤検出である可能性があるので、分割点尤度の更新は控えめに行う。

- [0099] このほかの特徴量についても同様に、その特徴量を抽出するのに必要十分な遅延時間を確保しつつ、分割点尤度特徴量抽出手段240と分割点尤度更新手段250を用いて、各発話の各ブロックに対する分割点尤度を更新する。
- [0100] 発話バッファリング手段230に保持されるあるブロックについて、そのブロックに対する分割点尤度更新が行われなくなったならば、分割点尤度決定手段260を用いてそのブロックを分割点として採用するかどうかを決定する。ただし、もし、一つの発話に対する分割数に上限を設けたり、その他の方法で分割前の発話全体に応じた処理を施すのであれば、特にいずれかの発話の終了時刻に相当するブロックに対する分割点尤度更新の完了を待たねばならない。
- [0101] なお、あるブロックに対する分割点尤度更新がどのタイミングで完了するかは、分割点尤度特徴量抽出手段240が要求する遅延時間と、分割点尤度更新手段250がその特徴量を反映させるブロックの範囲に応じて一意に決定できる。
- [0102] 分割点尤度決定手段260における分割点の決定と分割された発話の出力については、第2の実施の形態と同様である。
- [0103] 以上の説明では、便宜上、発話バッファリング手段230を二次元配列として表現したが、実際にはこの二次元配列の使われ方は疎である。従って、発話バッファリング手段230は、発話検出手段210で発話として検出された区間に相当するブロックのみを保持すればよい。
- [0104] この場合、分割点尤度決定手段260によって分割処理が完了したある発話について、その発話から分割点尤度特徴量抽出手段240によって得られる特徴量の影響を受ける他の発話の一つもないのであれば、その発話全体を発話バッファリング手段230から削除する。
- [0105] このようにして、必要最小限の記憶領域を用いて、オンラインに、本発明の原理に基づく発話分割が可能になる。
- [0106] 次に、本実施の形態の効果について説明する。
- [0107] 本実施の形態は、本発明による発話の分割を、最低限の遅延時間と必要最小限の記憶領域を用いて、オンラインに処理できる。従って本実施の形態は、本発明を非バッチ的なオンライン処理で用いる場合に好適な実施の形態であり、第1および第2の

実施の形態がバッチ的な用途に適しているのと対称をなす。

- [0108] 次に、本発明の第4の実施の形態について図面を参照して詳細に説明する。
- [0109] 図6を参照すると、本発明の第4の実施の形態は、図5に示した第3の実施の形態と比較して、音声認識手段330をさらに備える点で相違する。
- [0110] 音声認識手段330は、発話検出手段310によって切り出された発話ごとに、音声認識処理を施すことにより、単語列や単語ラティスなどの認識結果を出力する。
- [0111] その他の構成要素は第3の実施の形態とほぼ同様に動作するが、発話バッファリング手段340が発話情報の一つとして発話ごとの認識結果を保持する点と、分割点尤度特徴量抽出手段350および分割点尤度更新手段360が、発話の認識結果を分割点尤度特徴量として用いる点が異なる。
- [0112] 次に、図6を参照して本実施の形態の全体の動作について詳細に説明するが、本実施の形態の全体の動作は第3の実施の形態とほぼ同様であるため、以下では第3の実施の形態からの差分のみを記述する。
- [0113] 音声認識手段330を呼び出すタイミングは、音声認識手段330の機能的な制限によって二つのケースが考えられる。
- [0114] 第1のケースは、音声認識手段330が、発話の終端が確定しなくても順次その時点で確定した認識結果を出力でき、かつ、発話のオーバーラップする音声の数の上限が予め定められており、その上限に応じた音声認識手段330を複数同時に動作させるケースである。この場合は、発話検出手段310が、入力されたいずれかの音声から発話の開始を検出したタイミングから音声認識手段330が動作し、発話が継続する限り音声認識処理も継続され、認識結果が得られると、適宜発話バッファリング手段340に送る。
- [0115] 第2のケースは、音声認識手段330が、発話の終端が確定するまで音声認識結果を出力できない場合であるか、あるいは、潜在的に同時に起こり得るすべての発話に対して同時に複数の音声認識手段330を動作させられないか、のいずれかのケースである。この場合は、発話検出手段310が、入力されたいずれかの音声から発話の終端が検出されたタイミングで音声認識手段330を呼び出し、発話全体をまとめて音声認識して、その結果を発話バッファリング手段340に送る。

- [0116] いずれの場合も、分割点尤度特徴量抽出手段350および分割点尤度更新手段360の動作タイミングは若干の遅延が必要になるが、前者は、発話の一部に対する認識結果が確定するまでの時間だけの遅延となり、後者は、発話の終端が検出され、その発話全体に対する認識結果が得られるまでに掛かる時間だけの遅延となる。
- [0117] 分割点尤度特徴量抽出手段350は、いずれかの発話に関する認識結果が得られた時点で動作する。ここで、分割点尤度特徴量抽出手段350は、認識結果として得られた単語またはその単語の品詞を評価し、分割点尤度更新手段360は、予め単語および品詞ごとに定められた分割点尤度更新量に基づき、その認識結果が現れた時刻に相当する他の発話のブロックおよびその前後のブロックについて、分割点尤度を更新する。
- [0118] 例えば、「ええ」「ああ」「なるほど」「そうですね」などの、いわゆる「あいづち」を示す語彙が現れた場合、その時刻における他の発話の分割尤度は高くなる。
- [0119] 次に、本実施の形態の効果について説明する。
- [0120] 本実施の形態によれば、発話の分割において、その発話とオーバーラップする他の発話の語彙を参照することで、より精度の高い発話の分割を可能とする。これは、ある話者が意味的なあるまとまりの発話を完了した直後に、対話に参加する他の話者が発話する可能性の高い語彙や品詞に偏りがあるという性質を利用している。この性質は、例えば非特許文献2などで述べられている。
- [0121] [実施例]
- 次に、具体的な実施例を用いて本発明を実施するための最良の形態の動作を説明する。
- [0122] 図7に示すように、本発明の実施例は、音声入力デバイス410と、発話検出モジュール420と、音声認識モジュール430と、発話情報バッファ440と、単語アレイバッファ450と、分割尤度更新モジュール460と、認識結果分割モジュール470と、認識結果整列モジュール480と、表示装置490とから構成される。これらの各モジュールは、それぞれ概略次のように動作する。
- [0123] 音声入力デバイス410は、2チャンネルの音声信号をそれぞれ別々に入力して発話検出モジュール420に送る。

- [0124] 発話検出モジュール420は2つ用意されており、それぞれ音声入力デバイス410の各チャンネルに対応付けられている。
- [0125] それぞれの発話検出モジュール420は、入力された音声信号を順次解析し、発話の開始時刻および終了時刻を検出する。
- [0126] 音声認識モジュール430も2つ用意され、それぞれ異なる発話検出モジュール420に対応付けられている。
- [0127] それぞれの音声認識モジュール430は、対応する発話検出モジュール420が検出した発話に対して音声認識処理を施し、その発話区間の音声に対する音声認識結果を単語列として出力する。
- [0128] 発話情報バッファ440は、発話検出モジュール420が検出した発話ごとに、いずれのチャンネルから入力された音声であるかと、発話開始時刻と、発話終了時刻と、その認識結果単語列を記録する。また、各発話が処理済か否かを示すフラグも同時に格納される。
- [0129] 単語アレイバッファ450は、発話情報バッファ440からもっとも早い時刻に開始した未処理の発話を一つ取り出し、単語列を構成する単語と単語の間の個数だけの配列を確保する。この配列には、その単語と単語の間で発話を分割すべきかどうかを決定する分割点尤度が保持される。
- [0130] 分割点尤度更新モジュール460は、その時点で単語アレイバッファ450に展開されている分割処理対象発話について、時間的にオーバーラップする発話を発話情報バッファ440から検索する。見つかったオーバーラップする発話について、その開始時刻と、継続時間長と、処理対象発話とのオーバーラップ時間とを求め、これらに応じて単語アレイバッファ450の各要素である分割点尤度を更新する。また、オーバーラップ発話の認識結果単語列から、あいづちに相当する語彙を検索し、これが見つかった場合、この単語が発話された時刻にもっとも近い、単語アレイバッファ450上に展開されている要素(すなわち単語と単語の間)の分割点尤度を更新する。
- [0131] 認識結果分割モジュール470は、単語アレイバッファ450に展開されている発話について、その発話を構成する単語と単語の間における分割点尤度を参照し、予め定められた閾値を超えているならば、その位置を分割点として、発話を分割する。

- [0132] 認識結果整列モジュール480は、認識結果分割モジュール470によって分割された発話をその開始時刻の順に並べなおす。
- [0133] 表示装置490は、認識結果整列モジュール480の整列した順に、発話に相当する単語列を文字列として表示する。このとき、その発話がいずれのチャンネルから入力された音声に含まれるかをトレースし、そのチャンネル情報も併せて表示してもよい。また、その発話の開始時刻も併せて表示してもよい。なお、出力手段を構成する認識結果整列モジュール480は、認識結果を表示装置490に出力したが、図示しない記憶装置または後続する別のシステムやプログラムに出力するようにしても良い。
- [0134] 以上、本発明の実施の形態および実施例について説明したが、本発明は以上の実施の形態および実施例にのみ限定されず、その他各種の付加変更が可能である。また、本発明は、その有する機能をハードウェア的に実現することは勿論、コンピュータとプログラムとで実現することができる。プログラムは、磁気ディスクや半導体メモリ等のコンピュータ可読記録媒体に記録されて提供され、コンピュータの立ち上げ時などにコンピュータに読み取られ、そのコンピュータの動作を制御することにより、そのコンピュータを前述した各実施の形態および実施例における各手段として機能させる。
- [0135] この出願は、2008年2月19日に提出された日本出願特願2008-037200を基礎とする優先権を主張し、その開示の全てをここに取り込む。

産業上の利用可能性

- [0136] 本発明によれば、対話音声認識システム、とりわけ、特にインフォーマルな会議音声認識システムや、電話応対音声認識システムといった用途に適用できる。
- [0137] これらの対話音声認識システムで、特に対話音声認識結果を画面表示したり、あるいは記憶装置に出力する、対話テキスト化システム、会議録自動作成システム、電話応対記録システムなどに利用できる。
- [0138] また、音声対話の内容を解析する音声対話解析システムに適用することで、解析精度を向上させることが可能である。特に、各話者の発話内容の係り受けを解析する談話構造解析システムなどで有用である。

請求の範囲

- [1] 二人以上の話者による発話を入力とし、
それぞれの発話をさらに分割する発話分割手段を備え、
前記発話分割手段は、
ある発話の分割点の決定に際し、オーバーラップする他の発話の情報を用いること
を特徴とする発話分割システム。
- [2] 前記発話分割手段は、ある発話の分割点を求める際に、
オーバーラップする他の発話の開始時刻を少なくとも用いることを特徴とする請求
項1に記載の発話分割システム。
- [3] 前記発話分割手段は、ある発話の分割点を求める際に、
オーバーラップする他の発話の開始された時刻から分割点候補がどれだけ離れて
いるかを少なくとも用いることを特徴とする請求項1または2に記載の発話分割システ
ム。
- [4] 前記発話分割手段は、ある発話の分割点を求める際に、
オーバーラップする他の発話の継続時間長を少なくとも用いることを特徴とする請
求項1、2または3に記載の発話分割システム。
- [5] 前記発話分割手段は、ある発話の分割点を求める際に、
分割点候補にオーバーラップする他の発話の継続時間長を少なくとも用いることを
特徴とする請求項1、2、3または4に記載の発話分割システム。
- [6] 発話を入力とし、単語列等に変換する音声認識手段をさらに備え、
前記発話分割手段は、ある発話の分割点を求める際に、
分割点候補が現れた時刻またはその前後に、オーバーラップする他の発話の認識
結果に現れた語彙および品詞のいずれかまたは両方を少なくとも用いることを特徴と
する請求項1、2、3、4または5に記載の発話分割システム。
- [7] 二つ以上の音声のそれぞれに音声認識処理を施した結果を出力する複数音声認
識結果表示システムであって、
音声を入力とし、発話ごとに切り出す発話検出手段と、
発話を入力とし、テキスト等に変換する音声認識処理手段と、

- 前記音声認識手段に入力される発話を分割する請求項1乃至6の何れか1項に記載の発話分割システムとを備えたことを特徴とする複数音声認識結果表示システム。
- [8] 二つ以上の音声のそれぞれに音声認識処理を施した結果を出力する複数音声認識結果表示システムであって、
- 音声を入力とし、発話ごとに切り出す発話検出手段と、
- 発話を入力とし、テキスト等に変換する音声認識処理手段と、
- 前記音声認識処理手段により認識された音声認識結果を分割する請求項1乃至6の何れか1項に記載の発話分割システムと、
- 該発話分割システムの分割結果を出力する出力手段と、
- を備えたことを特徴とする複数音声認識結果表示システム。
- [9] 請求項7または8に記載の複数音声認識結果表示システムを使用して、二人以上の話者による対話音声を認識する対話音声テキスト化システム。
- [10] 請求項7または8に記載の複数音声認識結果表示システムを使用して、電話応対音声を音声認識する電話応対音声テキスト化システム。
- [11] 発話分割手段が、二人以上の話者による発話を入力して、それぞれの発話をさらに分割するステップを含み、
- 前記発話分割手段は、ある発話の分割点の決定に際し、オーバーラップする他の発話の情報を用いることを特徴とする発話分割方法。
- [12] 前記発話分割手段は、ある発話の分割点を求める際に、
- オーバーラップする他の発話の開始時刻を少なくとも用いることを特徴とする請求項11に記載の発話分割方法。
- [13] 前記発話分割手段は、ある発話の分割点を求める際に、
- オーバーラップする他の発話の開始された時刻から分割点候補がどれだけ離れているかを少なくとも用いることを特徴とする請求項11または12に記載の発話分割方法。
- [14] 前記発話分割手段は、ある発話の分割点を求める際に、
- オーバーラップする他の発話の継続時間長を少なくとも用いることを特徴とする請求項11、12または13に記載の発話分割方法。

- [15] 前記発話分割手段は、ある発話の分割点を求める際に、
分割点候補にオーバーラップする他の発話の継続時間長を少なくとも用いることを特徴とする請求項11、12、13または14に記載の発話分割方法。
- [16] 音声認識手段が、発話を入力とし、単語列等に変換するステップをさらに含み、
前記発話分割手段は、ある発話の分割点を求める際に、
分割点候補が現れた時刻またはその前後に、オーバーラップする他の発話の認識結果に現れた語彙および品詞のいずれかまたは両方を少なくとも用いることを特徴とする請求項11、12、13、14または15に記載の発話分割方法。
- [17] 二つ以上の音声のそれぞれに音声認識処理を施した結果を出力する複数音声認識結果表示方法であって、
発話検出手段が、音声を入力とし、発話ごとに切り出すステップと、
音声認識処理手段が、発話を入力とし、テキスト等に変換するステップと、
請求項1乃至6の何れか1項に記載の発話分割システムが、前記音声認識手段に入力される発話を分割するステップと、
を含むことを特徴とする複数音声認識結果表示方法。
- [18] 二つ以上の音声のそれぞれに音声認識処理を施した結果を出力する複数音声認識結果表示方法であって、
発話検出手段が、音声を入力とし、発話ごとに切り出すステップと、
音声認識処理手段が、発話を入力とし、テキスト等に変換するステップと、
請求項1乃至6の何れか1項に記載の発話分割システムが、前記音声認識処理手段により認識された音声認識結果を分割するステップと、
出力手段が、前記発話分割システムの分割結果を出力するステップと、
を含むことを特徴とする複数音声認識結果表示方法。
- [19] 請求項17または18に記載の複数音声認識結果表示方法を使用して、二人以上の話者による対話音声を認識する対話音声テキスト化方法。
- [20] 請求項17または18に記載の複数音声認識結果表示方法を使用して、電話対応音声を音声認識する電話対応音声テキスト化方法。
- [21] コンピュータを、

二人以上の話者による発話を入力とし、それぞれの発話を、その発話にオーバーラップする他の発話の情報を用いてさらに分割する発話分割手段として機能させるための発話分割プログラム。

- [22] 前記発話分割手段は、ある発話の分割点を求める際に、オーバーラップする他の発話の開始時刻を少なくとも用いることを特徴とする請求項21に記載の発話分割プログラム。
- [23] 前記発話分割手段は、ある発話の分割点を求める際に、オーバーラップする他の発話の開始された時刻から分割点候補がどれだけ離れているかを少なくとも用いることを特徴とする請求項21または22に記載の発話分割プログラム。
- [24] 前記発話分割手段は、ある発話の分割点を求める際に、オーバーラップする他の発話の継続時間長を少なくとも用いることを特徴とする請求項21、22または23に記載の発話分割プログラム。
- [25] 前記発話分割手段は、ある発話の分割点を求める際に、分割点候補にオーバーラップする他の発話の継続時間長を少なくとも用いることを特徴とする請求項21、22、23または24に記載の発話分割プログラム。
- [26] コンピュータを、さらに、発話を入力とし、単語列等に変換する音声認識手段として機能させるプログラムであって、
前記発話分割手段は、ある発話の分割点を求める際に、
分割点候補が現れた時刻またはその前後に、オーバーラップする他の発話の認識結果に現れた語彙および品詞のいずれかまたは両方を少なくとも用いることを特徴とする請求項21、22、23、24または25に記載の発話分割プログラム。
- [27] 二つ以上の音声のそれぞれに音声認識処理を施した結果を出力する複数音声認識結果表示システムを構成するコンピュータを、
音声を入力とし、発話ごとに切り出す発話検出手段と、
発話を入力とし、テキスト等に変換する音声認識処理手段と、
前記音声認識手段に入力される発話を分割する発話分割手段と、
して機能させるためのプログラムであって、

前記発話分割手段は、請求項1乃至6の何れか1項に記載の発話分割システムにより構成されていることを特徴とする複数音声認識結果表示プログラム。

- [28] 二つ以上の音声のそれぞれに音声認識処理を施した結果を出力する複数音声認識結果表示システムを構成するコンピュータを、
- 音声を入力とし、発話ごとに切り出す発話検出手段と、
- 発話を入力とし、テキスト等に変換する音声認識処理手段と、
- 前記音声認識処理手段により認識された音声認識結果を分割する発話分割手段と、
- 該発話分割手段の分割結果を出力する出力手段と、
- して機能させるためのプログラムであって、
- 前記発話分割手段は、請求項1乃至6の何れか1項に記載の発話分割システムにより構成されていることを特徴とする複数音声認識結果表示プログラム。

- [29] 二人以上の話者による対話音声を認識する対話音声テキスト化システムを構成するコンピュータを、
- 対話音声を入力とし、発話ごとに切り出す発話検出手段と、
- 発話を入力とし、テキスト等に変換する音声認識処理手段と、
- 前記音声認識手段に入力される発話を分割する発話分割手段と、
- して機能させるためのプログラムであって、
- 前記発話分割手段は、請求項1乃至6の何れか1項に記載の発話分割システムにより構成されていることを特徴とする対話音声テキスト化プログラム。

- [30] 二人以上の話者による対話音声を認識する対話音声テキスト化システムを構成するコンピュータを、
- 対話音声を入力とし、発話ごとに切り出す発話検出手段と、
- 発話を入力とし、テキスト等に変換する音声認識処理手段と、
- 前記音声認識処理手段により認識された音声認識結果を分割する発話分割手段と、
- 該発話分割手段の分割結果を出力する出力手段と、
- して機能させるためのプログラムであって、

前記発話分割手段は、請求項1乃至6の何れか1項に記載の発話分割システムにより構成されていることを特徴とする対話音声テキスト化プログラム。

[31] 電話応対音声を音声認識する電話応対音声テキスト化システムを構成するコンピュータを、

電話応対音声を入力とし、発話ごとに切り出す発話検出手段と、
発話を入力とし、テキスト等に変換する音声認識処理手段と、
前記音声認識手段に入力される発話を分割する発話分割手段と、
して機能させるためのプログラムであって、

前記発話分割手段は、請求項1乃至6の何れか1項に記載の発話分割システムにより構成されていることを特徴とする電話応対音声テキスト化プログラム。

[32] 電話応対音声を音声認識する電話応対音声テキスト化システムを構成するコンピュータを、

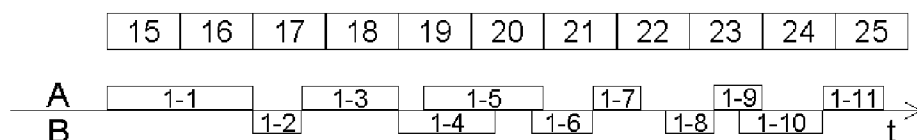
電話応対音声を入力とし、発話ごとに切り出す発話検出手段と、
発話を入力とし、テキスト等に変換する音声認識処理手段と、
前記音声認識処理手段により認識された音声認識結果を分割する発話分割手段と、

該発話分割手段の分割結果を出力する出力手段と、
して機能させるためのプログラムであって、

前記発話分割手段は、請求項1乃至6の何れか1項に記載の発話分割システムにより構成されていることを特徴とする電話応対音声テキスト化プログラム。

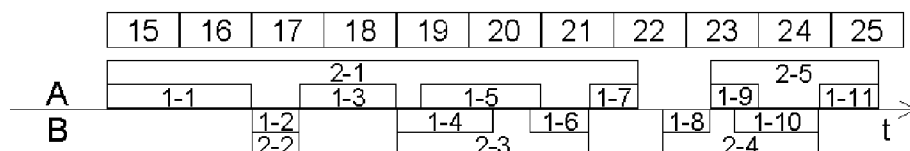
[図1]

発話番号	時刻	話者	発話内容
1-1	20:30:15.0	A	先ほどお伺いした件なんですが...
1-2	20:30:17.0	B	はい。
1-3	20:30:17.8	A	...まだご連絡...
1-4	20:30:19.0	B	あ、ご連絡まだ...
1-5	20:30:19.4	A	...頂けてないんですけども。
1-6	20:30:20.7	B	...ですか。
1-7	20:30:21.5	A	はい。
1-8	20:30:22.8	B	それは...
1-9	20:30:23.3	A	ええ...
1-10	20:30:23.5	B	... 1時ごろの。
1-11	20:30:25.2	A	いえ、3時ごろの。

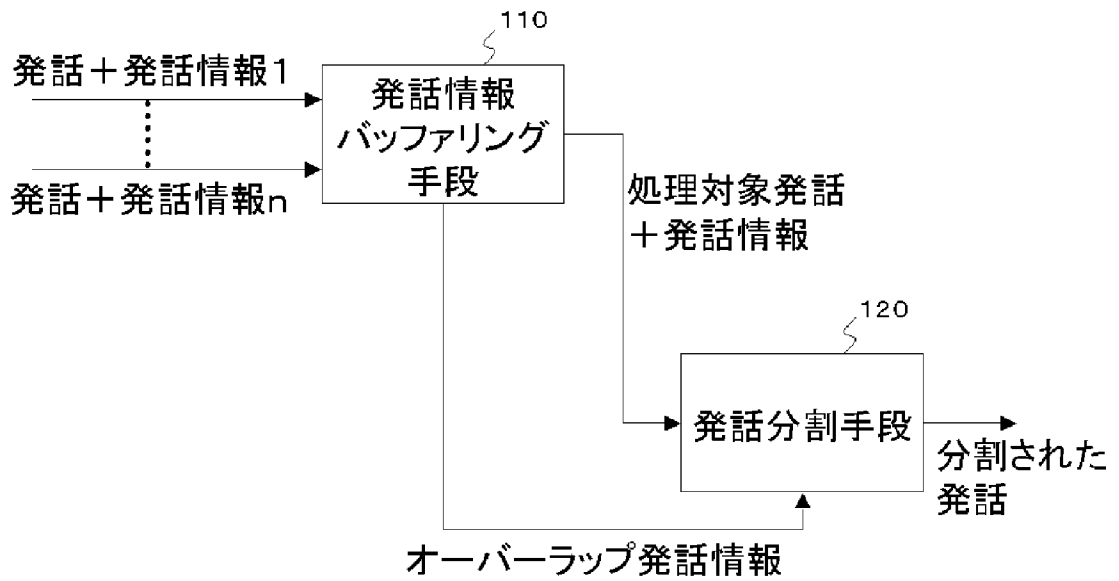


[図2]

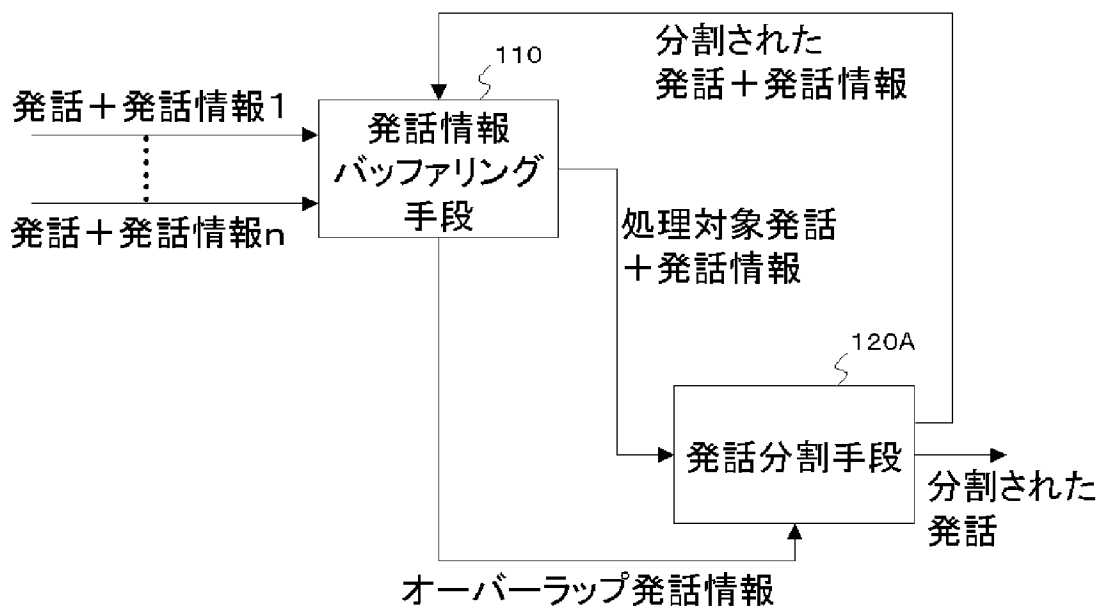
発話番号	時刻	話者	発話内容
2-1	20:30:15.0	A	先ほどお伺いした件なんですが、 まだご連絡頂けてないんですけども。 はい。
2-2	20:30:17.0	B	はい。
2-3	20:30:19.0	B	あ、ご連絡まだですか。
2-4	20:30:22.8	B	それは1時ごろの。
2-5	20:30:23.3	A	ええ...。いえ、3時ごろの。



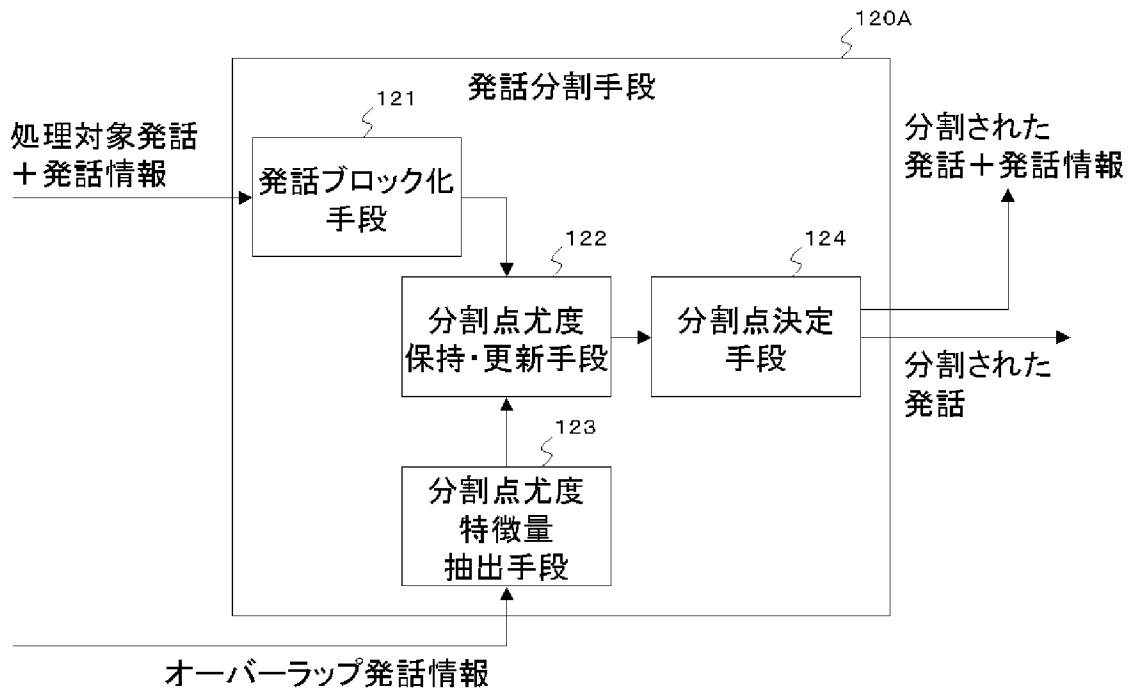
[図3]



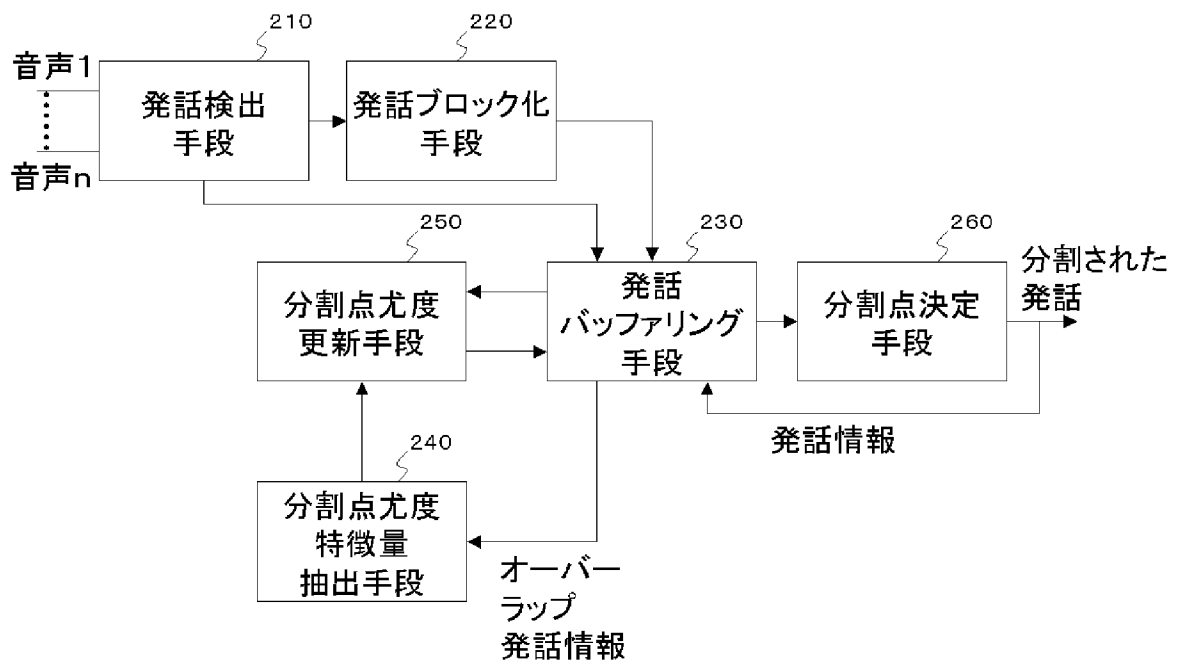
[図4A]



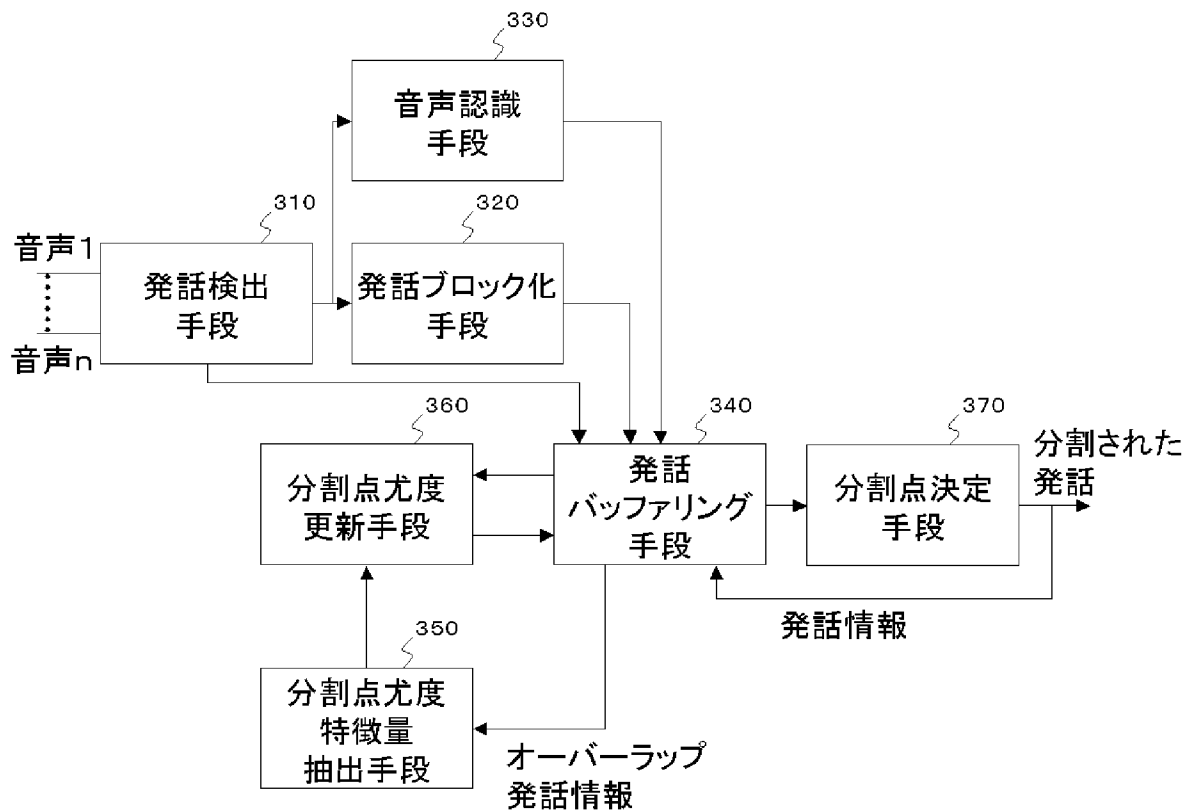
[図4B]



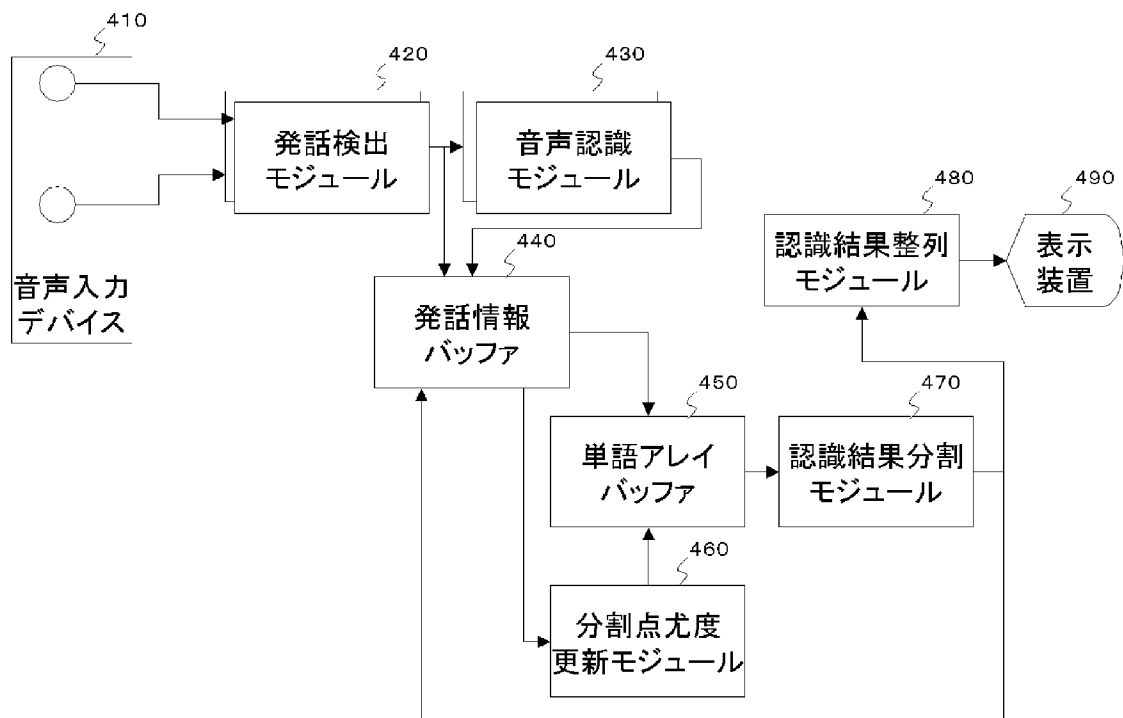
[図5]



[図6]



[図7]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2008/072851

A. CLASSIFICATION OF SUBJECT MATTER
G10L15/04(2006.01)i, G10L15/00(2006.01)i, G10L15/22(2006.01)i, G10L15/28(2006.01)i, G10L17/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L15/00-17/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2009
Kokai Jitsuyo Shinan Koho	1971-2009	Toroku Jitsuyo Shinan Koho	1994-2009

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPI

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2005-202035 A (Toshiba Corp.), 28 July, 2005 (28.07.05), Full text; all drawings (Family: none)	1-5, 11-15, 21-25
Y	JP 2005-308950 A (Sony Corp.), 04 November, 2005 (04.11.05), Full text; all drawings (Family: none)	6-10, 16-20, 26-32
Y	JP 2005-308950 A (Sony Corp.), 04 November, 2005 (04.11.05), Full text; all drawings (Family: none)	6-10, 16-20, 26-32
A	JP 10-69292 A (ATR Interpreting Telecommunications Research Laboratories), 10 March, 1998 (10.03.98), Full text; all drawings (Family: none)	1-32

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 23 March, 2009 (23.03.09)	Date of mailing of the international search report 31 March, 2009 (31.03.09)
--	---

Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/072851

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2002-236494 A (Denso Corp.), 23 August, 2002 (23.08.02), Full text; all drawings (Family: none)	1-32
A	JP 11-136369 A (NTT Software Corp.), 21 May, 1999 (21.05.99), Full text; all drawings (Family: none)	1-32

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G10L15/04(2006.01)i, G10L15/00(2006.01)i, G10L15/22(2006.01)i, G10L15/28(2006.01)i, G10L17/00(2006.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int.Cl. G10L15/00-17/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2009年
日本国実用新案登録公報	1996-2009年
日本国登録実用新案公報	1994-2009年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

WPI

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP 2005-202035 A (株式会社東芝) 2005.07.28, 全文, 全図 (ファミリーなし)	1- 5, 11-15, 21-25
Y		6-10, 16-20, 26-32
Y	JP 2005-308950 A (ソニー株式会社) 2005.11.04, 全文, 全図 (ファミリーなし)	6-10, 16-20, 26-32

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的な技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

23.03.2009

国際調査報告の発送日

31.03.2009

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)
郵便番号100-8915
東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

山下 剛史

5Z

8946

電話番号 03-3581-1101 内線 3541

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 10-69292 A (株式会社エイ・ティ・アール音声翻訳通信研究所) 1998.03.10, 全文, 全図 (ファミリーなし)	1-32
A	JP 2002-236494 A (株式会社デンソー) 2002.08.23, 全文, 全図 (ファミリーなし)	1-32
A	JP 11-136369 A (エヌ・ティ・ティ・ソフトウェア株式会社) 1999.05.21, 全文, 全図 (ファミリーなし)	1-32