

車載情報機器の音声操作における文脈の保持とタイムアウト制御

木下悠[†] 川端豪[†]

車載情報機器のハンズフリー操作のために音声インタフェースが有望である。さらにその設計においては運転中の緊急の妨害を想定することが必要である。ナビゲーションシステムによく用いられている木構造のメニューを順にたどる操作系があるが、この枠組は音声で項目を順にたどる方式に容易に拡張でき、また妨害による中断に対応できる。各項目は木構造中の節点となり、その親節点が指定されている。例えば、「気温」というコマンド発話を行うためには、先行して「天気」というコマンド発話が必要である。ある親節点が受理されると、子節点の項目が活性化され、対応するコマンドが音声認識の辞書に加えられる。本報告ではこの親節点のことを「文脈」と表現する。各文脈には有効時間が与えられ、しばらく沈黙が続くと無効化される(タイムアウト)。このタイムアウトの設定に関して、種々の検討を行った。例えば、コマンド階層間における適切なタイムアウト値の設定。また発話中に妨害が入って中断が起こった場合のタイムアウト値の設定などについて述べる。

Timeout Controls of the Voice Command Context for Telematics Systems

YU KINOSHITA[†] TAKESHI KAWABATA[†]

Speech interface is promising for hands-free operations of telematics systems. For designing information devices in the car, it is necessary to take emergent interruption into account. The tree-based item selection method, which is used in navigation systems, can be applicable to speech command interfaces. A menu item assigned to a task function is a node that has a parent node. For example, the item "Show the air temperature" has the parent item "Weather". The user has to pronounce "Information", "Weather" and "Air temperature" consequently. In this case, we call the parent node as the voice command context. The context activates its child items as the speech recognition vocabulary. The system gives a proper time to each context. We call it the timeout. This paper reports miscellaneous approach for the timeout control. For example, we investigate the timeout taking the time period between the layers into account. And we investigate the appropriate timeout when the spoken word is interrupted.

1. 序論

無線通信ネットワークの整備が進み、車載情報機器の操作が複雑化するにつれ、そのヒューマンインタフェース設計の重要性が高まってきている。中でもハンズフリーで機器の操作を行うことができる音声インタフェースは、利便性、安全性の観点から注目されている[1]。篠原らは運転中の情報機器利用における肯定的及び否定的影響についてまとめている[2]。町田らは注意の分散がもたらす潜在的な危険性についても報告されている[3]。また、評価に関する種々の研究が行われている[4][5][6]。車載情報機器を音声操作において、伊藤らは目的地設定をタスクとした対話状況の違いによる音響的特徴を報告している[7]。

自動車運転中に情報機器を操作する場合、操作中断の問題を避けて通ることができない。例えば、ナビゲーションを音声操作している途中で、路上の危険を発見し最優先で危険回避行動を行った結果、ナビゲーション操作は中断する。危険回避後に、始めから機器操作をやり直すか、途中から再開できるかでインタフェースの操作性は大きく変わってくる。一度有効になったサービスの状態を、どのくらいの時間保持するか(タイムアウト)の設計が必要になる。

対話型インタフェースにおいて、タイムアウトの概念は極めて重要で、ユーザーが意識するしなないに関わらず何らかの形で組み込まれていることが多い。

車載情報機器のメニュー操作において、コマンド項目を階層化された木構造中に配置することがよく行われる。図1にナビゲーションシステムの典型的なコマンドツリー例を示す。

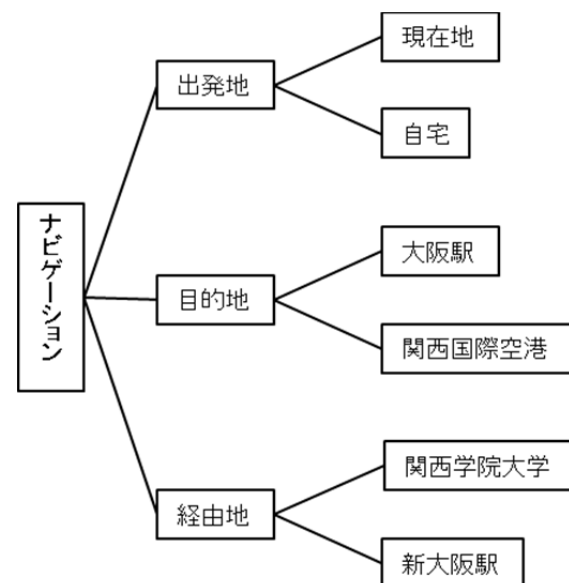


図1 ナビゲーションのコマンドツリー例

[†] 関西学院大学
Kwansei Gakuin University

タッチパネルにて「ナビゲーション」、「出発地」、「自宅」と順に項目を選択していくことで目的の操作が達成される。同様の選択を音声インタフェースによって行う場合には、「ナビゲーション」、「出発地」、「自宅」という単語を順に発声することで同じ操作が行える。木構造の節点に配置された項目が受理されるためには、それに先行してその親節点の項目が発声され受理されていなければいけない。すなわち各項目には木構造によって指定される親子構造が与えられている。

言い換えれば、例えば「出発地」という節点が受理された状態で、その子である「現在地」及び「自宅」という単語が音声認識の対象として有効化（アクティベーション）していると考えることができる。ここでその親節点のことを「文脈」と呼ぶことにする。その文脈が保持される時間の設定には色々なやり方が考えられ、音声インタフェースの操作性に影響を及ぼすと考えられる。ある文脈に対する有効時間のことを本報告ではタイムアウトと呼び、その設定方法を検討する。

音声で車載情報機器のメニュー階層をたどる際に、発話中断の影響を考慮した仕組みを構築する。また、下の階層へのコマンド連続発話と発話中断前の文脈の保持を可能にし、その文脈のタイムアウトを定める。

メニューコマンドは階層構造になっており、上の階層の語彙が発声されると、その語彙に対応した下の階層の語彙が有効化される。その上階層の文脈は発話中断後も適切な時間保持されるべきであり、タイムアウトの設定には様々な考え方があり検討する必要がある。また対話において、人は対話の流れを考えて（プランニング）音声操作を行うため、中断によってプランが壊れた際の状況も検討する。

2. 文脈の保持とタイムアウト制御を考慮した音声操作

2.1 目的と方法論

カーナビゲーションによく用いられているような、階層的なメニューを順にたどる操作を、音声コマンドによって実現する際、走行状況に基づく発話中断などがあっても快適に操作が行えるような枠組みを構築する。

深い階層をたどることを想定し、コマンド語彙を連続して発話できるようにする。発話中断後に続けて階層メニューをたどれるように中断前の文脈を保持する。このための各文脈の選択項目に対する有効時間（タイムアウト）を設定する。

実際の情報システムを模した実験システムを被験者に使用してもらい、種々の発話のタイミングに関する時間測定を行う。平均値や分布を観察し、タイムアウト設定に関する知見を得る。

2.2 文脈による語彙のアクティベーション

階層的項目選択によるコマンド絞込みの方式について図2を用いて説明する。コマンドはいくつかの階層に分類して配置されている。階層0には「機器操作」という語彙が配置されており、音声コマンドの開始に際し、まずこの語彙を発声しなければならない。階層1には例えば、「ナビゲーション操作」、「情報検索」、「オーディオ操作」といった語彙が配置され、車載コンピューターに与える指示の大分類を担う。続いて階層2以降に各分類の操作の種類が配置され、このようなツリー構造をたどることにより、最終的に目的とする操作を選択する。図2は説明のために部分的に表示しているが、コマンドツリーの全体図を報告の末尾に補足資料として挙げておく。

図2のメニュー項目をたどる操作において、深い階層にある項目を選択するために、コマンド語彙を連続して発話できるようにすることが必要である。一気に必要な語彙を連続発話することもあれば、語彙を一つずつ発声し、順に階層をたどることも許容する。さらに、発話の途中で危険回避が必要になりコマンドが中断した場合、復帰後に途中の階層から始められるような枠組みが望まれる。

上記のような機能を実現するために、本報告では文脈の保持とタイムアウト制御に基づく手法を考案する。例えば、図2において、まずユーザーは、「機器操作」と発声する。これによって機器操作という文脈が選ばれ、それに対応した語彙の集合が音声認識の対象として有効化（アクティベーション）される。図中では「ナビゲーション操作」、「情報検索」、「オーディオ操作」が有効化される。ユーザーがそのどれかの語彙を発声すると、その選択が新しい文脈となり、さらに下の階層の対応する語彙集合が有効化される。これらの文脈の有効化の連鎖はリアルタイムに行われるため、ユーザーは語彙の連鎖を連続的に発声してもよい。

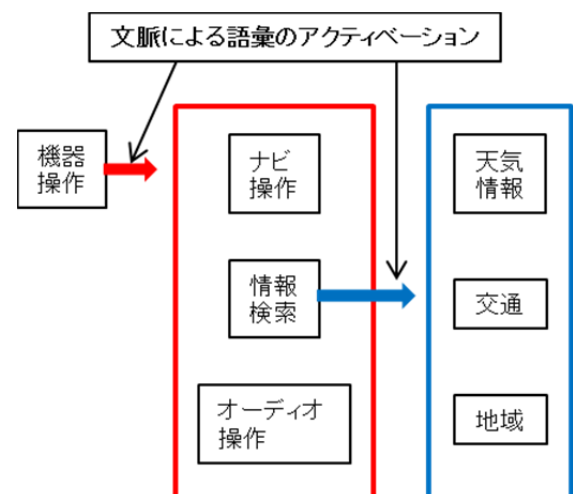


図2 階層的項目選択における文脈と語彙の有効化

発話中断が起きた場合は、その時点で有効化されている文脈はまだしばらくの間保持される。この時間長をタイムアウトと呼ぶ。図3に様々なタイムアウト設定の考え方を示す。(a)に示す無限大型は、タイムアウトを無限大に設定することで、一度たどられた文脈については全ての語彙を再発声することが可能になる。これは中断からの復帰においては便利であるが、多数の語彙が認識対象として同時に有効になるので、誤認識の可能性も高まる。一方で、(b)の次発話型は子である語彙を発声すると、その親はタイムアウトする。このようにタイムアウトを小さく設定すると、中断があった場合にその文脈がすぐにタイムアウトしてしまっているので、途中からの再発声ができなくなる。(c)の更新型は、最後に発声された語彙の階層よりも上の階層で発声された文脈のタイムアウトを更新する。そのことによって、一番下の階層から順にタイムアウトしていく形になり、適切なタイムアウト設定を行うことが重要だと考えられる。またユーザーは対話が始まる前に対話のシナリオを考えている(プランニング)と予想できる。そのため発話中断が起きた際、そのプランが壊され、一からもしくは途中から再プランニングする必要がある。その再プランニングの時間は、発話中断後から次の発話までの時間と考えられるので、発話中断が起きた場合にはその時間も考慮に入れる必要がある。

3. タイムアウト決定のための実験的検討

前節までに述べてきたように、階層型メニューを音声コマンドでたどる操作において、各文脈のタイムアウトの設定が重要である。本節では、このタイムアウトすなわちある文脈が保持される時間の設定について検討する。

3.1 実験方法

被験者は20代男性4名、女性1名とした。被験者には実験を始める前にタスクや発声方法について十分に理解してもらい、また本番のタスクに似た練習課題をいくつかこなしてもらった。本番ではタスクを達成するまで何度も言い直してもらうようにした。本実験は図3の更新型のタイムアウトを決定することも目的の一つであるため、今回は無限大型のタイムアウトで実験を行った。

本研究は車載機との音声インタフェースの情報検索、ナビゲーション操作、オーディオ操作を想定しているため、それらについてのタスクを用意した。用意されたタスクの内容を検索・設定・操作できればタスク達成とする。タスクの具体例を下記に示す。

- 今日の天気、平均気温、降水量を調べて下さい
- ナビの新規設定で出発地を自宅、目的地を関学、経路地を有馬、経路を経路2に設定して下さい

- “アーティスト名”の“曲名”を再生し、音量を5くらい上げて下さい

システムとの対話は単語発話によって行われる。具体例の一つ目においては、「機器操作」「情報検索」「天気情報」「今日」「天気」「平均気温」「降水量」と順番に発話することでタスク達成ができる。ユーザーの1つの単語に対して、システムは1つの応答をするが、ユーザーは単発単語発話だけでなく、連続単語発話が可能である。システムとの対話例を図4に示す。また発話中断を起こすために、被験者には妨害のタイミングを伝えず、ある決まったタスクの単語をユーザーが発話した後に2桁×1桁の掛け算を暗算で解いてもらった。これは発話中断前のプランニングは壊すことが目的である。

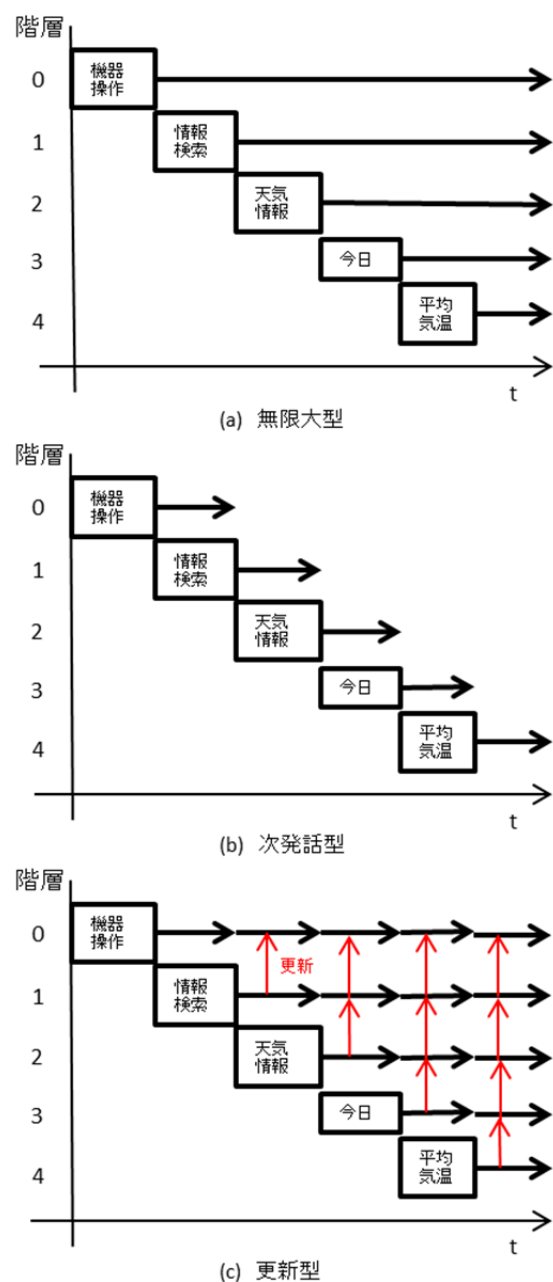
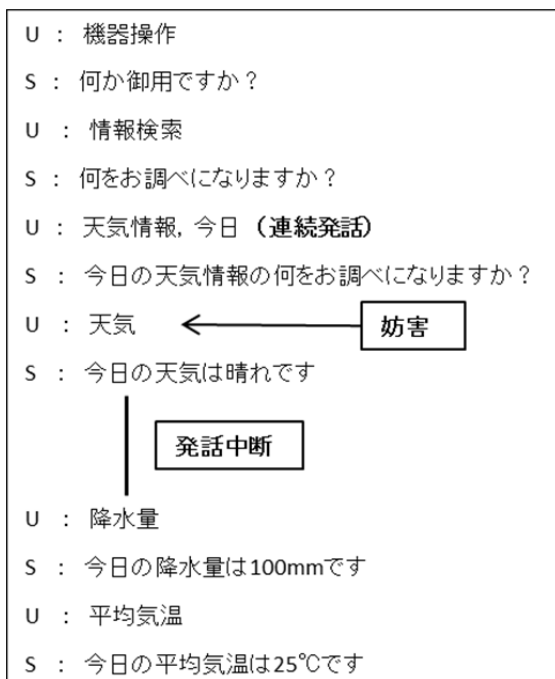


図3 文脈(発話項目)とその有効区間(タイムアウト)

3.2 実験結果と考察

図5に階層間におけるタイムアウトの平均値を示す。システムの発話後からユーザーの次の発話が丁度認識されるまでを測定したため、妨害や誤認識がない場合はこの平均値よりもプラス2秒か3秒を各階層のタイムアウトとするべきだと考えられる。0層と1層の間のタイムアウトが一番長く、下の階層に進むに従ってタイムアウト時間は短くなっている。これは上の階層ではそれ以降の階層のプランニングを行うため、上の階層ほどプランニングの量が多くなる。逆に下の階層ほどそれ以降のプランニングが少なくなるため、図5のような結果になったと考えられる。本実験の4層構造の場合、0層1層の間と3層4層の間のタイムアウトは約1秒の差が生まれた。

図6に階層ごとの発話中断後からの復帰時間の平均値を示す。3層目までは階層が深くなるごとに復帰時間が長くなっている。階層が深くなるにつれてユーザーの記憶すべき文脈の数が多くなる。そのため発話中断において、文脈を思い出す時間が長くなったことと、階層の途中では次の発話のプランニングもしなければいけないことから、図6のような結果になったと考えられる。よって深い階層ほど妨害の再プランニングへの影響は大きくなると考えられる。しかし、最後の階層では復帰時間が少し短くなったため、最終階層については別の考えが必要である。最終階層は目的とする選択の終着点であることから、次へのプランニングの量がとても少ない。そのことから図6の4層では少し復帰時間が下がったと考えられる。



U : ユーザー S : システム

図4 音声対話システムとの対話例

図7に連続発話における発話間隔時間の分布を示す。図から0.2秒前後の時間に分布の中心があることがわかる。連続発話に先行して、ユーザーがプランニングをしていると考え、この結果から、プランニングを伴う発話における単語間隔は0.2秒前後であると考えられる。連続発話個数は4個が5回, 3個が12回, 2個が35回となった。ユーザーにとって、最も必要な連続発話個数は2個であると推測することができる。

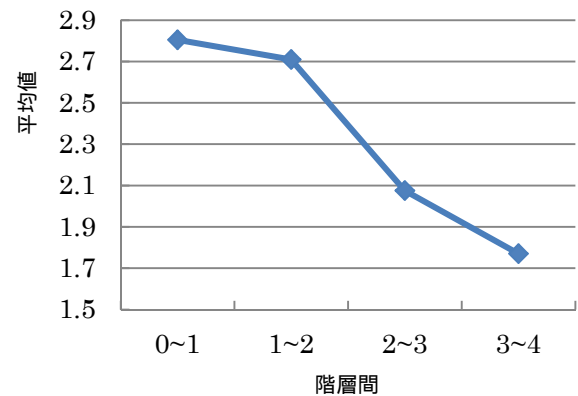


図5 階層間ごとのタイムアウト平均値

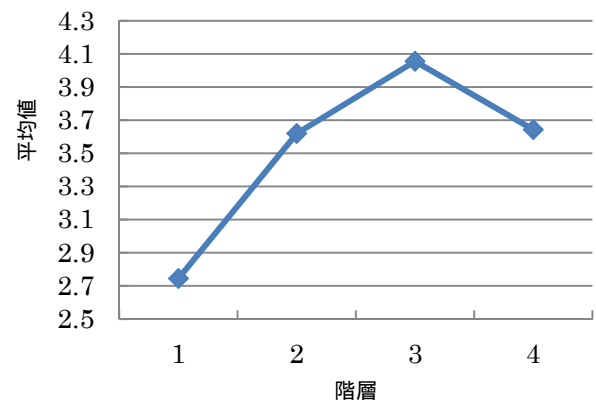


図6 階層ごとの発話中断後からの復帰時間の平均値

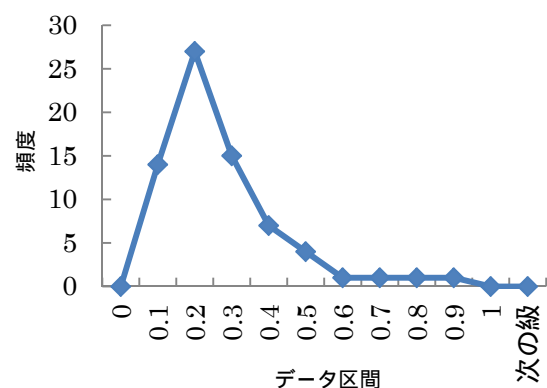


図7 連続発話における発話間隔時間の分布

4. 結論

4.1 まとめ

音声操作において、階層メニューの項目をたどるときの文脈のタイムアウトについて実験的検討を行った。システムの発話からユーザーの次発話までの時間を測定した結果、階層間ごとのタイムアウトは階層が深くなるにつれて小さくなることがわかった。

また、発話中に妨害を入れることで発話中断を起こし、中断後から次の発話に復帰するまでの時間も測定した。その結果から、深い階層に進むに従って復帰時間は長くなった。しかし、最終階層は復帰時間が少し短くなり、他の階層とは違うことが示唆された。

ユーザーのプランニングの際の発話間隔を調べるために、連続発話された単語の間隔を測定した。発話間隔時間は0.2秒前後に多く分布した。

4.2 今後の課題

今回の実験では被験者が5人と少なかったため、被験者を増やしていき、より確実な結果にする。また、システムに慣れてからでないと連続発話ができない傾向が見られたので、事前の練習についても見直す。

実験結果から、更新型の適切なタイムアウトがわかったため、無限大型、次発話型、更新型の使い比べを行い、ご認識率の測定やユーザーの再発話数、3つのシステム主観的評価を行っていく。

参考文献

- 1) 脇田:「自動車における音声インタフェース」、音学誌, Vol.59, No.2, pp.94-98 (2003)
- 2) 篠原:「情報機器利用による自動車運転者の注意への肯定的影響と否定的影響」、大阪大学大学院人間科学研究科紀要, Vol.30, pp.15-34(2004)
- 3) 町田:「自動車運転時のハンズフリーシステムの有効性に関する研究」、日本人間工学会大会講演集,(2010)
- 4) 清水:「運転中における音声対話システムの評価」、情報処理, pp.87-92(2000)
- 5) 由浅:「状況と文脈を利用した音声対話型車載インタフェースの構築と評価」、情報処理, 2003(124), pp.199-204
- 6) 駒谷:「ユーザモデルを導入したバス運行情報案内システムの実験的評価」、情報処理, 2003(75), pp.59-64
- 7) 伊藤:「目的地設定タスクにおける対話状況の違いによる言語・音響的特徴」、情報処理, Vol.43, No.7, pp.2118-2129

補足資料

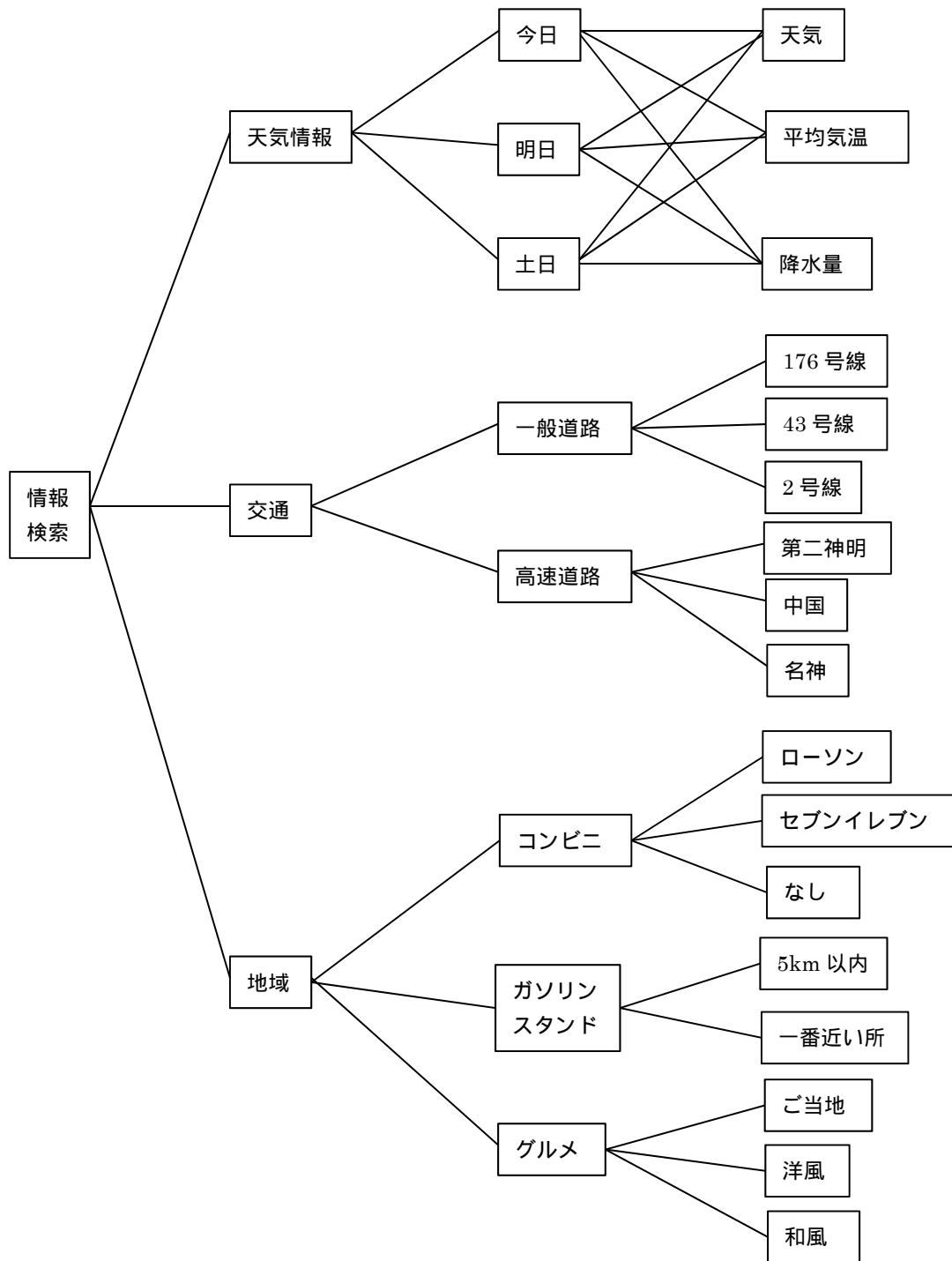


図8 情報検索ツリーの全体図