

トリガベース音声対話システムのアーキテクチャ提案

Proposal for the architecture of trigger based voice interactive system

段 清柱† 関口 隆昭† 清水 淳史†
Qingzhu Duan Takaaki Sekiguchi Atsushi Shimizu

1. はじめに

近年の音声認識技術の向上に伴い、Siri^{*1}やしゃべってコンシェル^{*2}等、ユーザの自然な発話の理解やユーザとの自然な対話（音声対話）による情報検索サービスが実用レベルになっている[1]。音声対話による情報検索は、特に自動車の運転中におけるカーナビの目的地設定等、手で操作することが難しい場面での応用が期待されている。

音声対話システムに関する従来の研究は、テキスト入力による Web 検索と同様、ユーザの質問に対して情報を検索して回答するユーザ主導型のシステムが多い。このようなユーザ主導型のシステムでは、どのような発話文であれば適切な回答を得られるかをユーザがよく考えながら質問する必要がある。ユーザが対話処理に集中している場合は問題ないが、車内の場合、ユーザは音声対話ではなく車両の運転に集中しており、運転しながら適切な発話文を考えて対話を継続することは難しい。よって、車内で使い易い音声対話サービスを提供するためには、ユーザ主導型に加え、ユーザの状況に応じて情報提供を行うシステム主導型の音声対話を実現する必要がある。

以上の背景を踏まえ、本稿では、車内においてシステム主導型の音声対話を実現するシステム（以降、トリガベース音声対話システムと呼ぶ）を提案する。

2. 関連研究

文献[2]では、観光地の紹介を目的として、ユーザ主導とシステム主導の2つのモードを切り替えて対話を継続するシステムについて述べている。このシステムでは、最初はユーザ主導の質問に対して観光地の情報提供を行うが、対話中に数秒間の沈黙を検出した場合はシステム主導に切り替わり、ユーザの興味がありそうな情報を推薦する。

この手法を車内に応用する場合、以下が問題になると考える。

(1) システム主導発話を行うための情報収集

観光地の紹介のように音声対話用の端末が特定の場所に固定化されている場合と異なり、車内の場合はユーザの興味は車両の現在位置や周辺状況等により変化する。そのため、これらの情報を収集してシステム主導発話を開始する方式について明らかにする必要がある。

(2) システム主導発話のタイミング制御

システム主導で発話するタイミングは、沈黙の検出の他に様々な判断方法が考えられる。文献[2]でも、沈黙以外の情報からユーザの心的状態を察知する必要性を言及しており、特に車内の場合は、車両周辺の状況等によりユーザが発話に回答できない場合があり、考慮が必要である。

3. トリガベース音声対話システムの提案

3.1 提案システム構成

図1に、本稿で提案するトリガベース音声対話システムの構成を示す。音声対話に関する機能構成や処理手順は従来のシステムと同様であり、以下の流れで実行する。

- (1) 車載端末から音声対話サーバに対して、ユーザが発話した音声を送信する。
- (2) 音声認識機能によって音声を変換する。
- (3) 意図解釈機能によって上記テキストからユーザが実行したい内容（意図）を判断する。
- (4) ユーザの意図に従い必要な情報をコンテンツから検索して応答文章を生成する。
- (5) 音声合成機能によって応答文章を音声に変換する。
- (6) 音声対話サーバから車載端末に音声を送信し、ユーザに回答音声を伝える。

以上の機能に加え、提案システムでは、2章で挙げた問題点を解決するためのトリガ管理機能を設ける。トリガ管理機能は、システム主導発話のきっかけとなり得る情報（これを0次トリガと呼ぶ）を入力として受け、対話管理機能にシステム主導発話の開始を通知（これを1次トリガと呼ぶ）する機能である。以降で詳細を述べる。

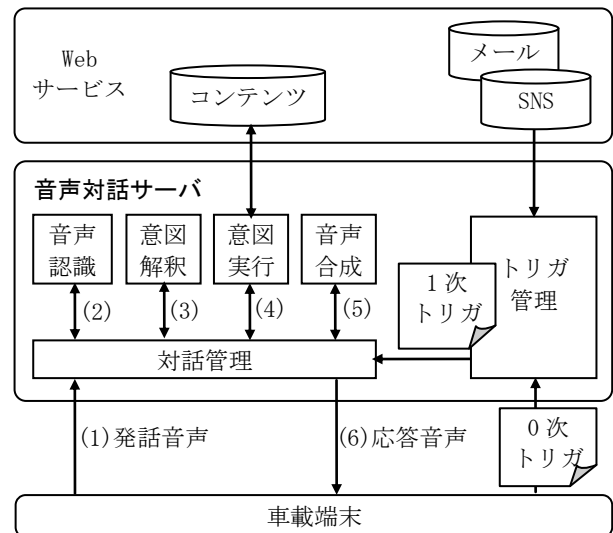


図1. トリガベース音声対話システムの構成

3.2 システム主導発話を行うための情報収集

表1は、ユーザの状況を示す情報について、情報の発生源と、情報が変化する頻度によって分類したものである。項番1は、車載端末で発生する動的な情報であり、例えばカーナビの目的地、車両の現在地や速度、電源ポジション（ACC等）、車両情報（CAN情報）である。項番2は、

† (株) 日立製作所 横浜研究所

*1 Siri は米国 Apple Inc. の登録商標

*2 しゃべってコンシェルは NTT ドコモ社の登録商標

車載端末で発生する静的な情報であり、例えばユーザの嗜好を分析した結果等である。項番3は提案システムの外部のWebサービスから取得可能な動的な情報であり、例えばメールやニュースの受信状態である。最後の項番4は、Webサービスから取得可能な静的な情報であり、例えばTwitter等のSNS情報の分析結果(流行情報)等である。

表1. ユーザの状況を表す要素

項番	情報源	頻度	例
1	車載端末	動的	ナビ情報(目的地設定等) 走行状態(位置、車速等) 車両状態(電源ポジション等) 車両情報(CAN情報等)
2		静的	個人属性(操作履歴、嗜好等)
3	Web	動的	メール/ニュース受信状態等
4	サービス	静的	SNS分析(流行情報)等

これらの情報を参照してシステム主導で発話を行う方式として、本稿では情報の変化をトリガにして発話を開始するルールベースの方式を提案する。トリガとして、表1に示す情報のうち動的に変化する情報(項番1、項番3)が考えられるが、項番3はWebサービスによって提供される情報であり、一般的なRESTモデルのWebAPIを用いてポーリング等により各サービスに定期的にアクセスすることを考慮すると、情報の変化に応じてリアルタイムに発話を行うトリガにはなりにくい。そこで項番1の情報の発生(0次トリガ)を起点に、項番2から項番4の情報を参照してルールに従い、システム主導対話開始の通知である1次トリガを生成する方式とする。

3.3 システム主導発話のタイミング制御

2章で挙げた問題点(2)について、システム主導の発話を行うタイミングをどのように決定するかを以下に記載する。

図2に、トリガ管理機能の構成を示す。

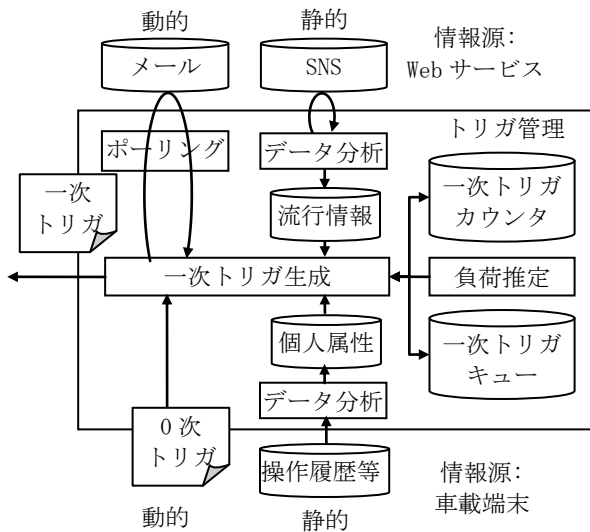


図2. トリガ管理機能の構成

3.2節で述べたように、提案方式は0次トリガの発生をきっかけにルールベースでシステム主導発話を開始する。その場合、同一の0次トリガに対して何度も同じ発話が行われる可能性があり、また、0次トリガの発生に対して無条件で発話を行うと、交差点等の運転操作に集中すべき場所ではユーザの運転操作の妨げになる可能性がある。

そのため提案方式では、同じ発話が繰り返されることを抑止するため、同一の0次トリガの発生を記録し、この発生回数が閾値を超えた場合には一次トリガの生成を抑止する機構を設ける(これを一次トリガカウンターと呼ぶ)。また、無条件で一次トリガが生成されることを抑止するため、0次トリガの情報からユーザの負荷を推定し、負荷が高い場合には生成した一次トリガを一時的に蓄積する機構(これを一次トリガキューと呼ぶ)を設ける。

4. 考察

3章で述べた方式によるシステム主導発話の有効性について、以下に考察を記載する。

(1) 拡張性

提案方式は、0次トリガに基づいてルールベースでシステム主導発話を開始する。一般的にルールベースで情報をレコメンドする方式は、ルールを設定する事業者の意思を反映できるメリットがある一方、ルール数が多くなると運用コストが増大するというデメリットがある。ルール数に関して、車載端末で取得できる設定情報や走行状態(位置、車速、加速度等)は高々数十種類であり問題ないと思われるが、車両から取得できるCAN情報(アクセル、ブレーキ、舵角等)の場合は無数にあり、これらも0次トリガとして扱う場合、他のレコメンド技術(協調フィルタリング等)の応用を検討する必要があると考えられる。

(2) 利便性

提案方式は、システム主導発話のタイミングを、車載端末からの情報(0次トリガ)に基づいて簡易的に判断している。これは方式の単純さ、わかりやすさを優先したためであるが、ユーザの負荷を推定する方式については人間工学に基づく精神的作業負荷を定めたISO10075等、各種の取組があり、これらの先行的取組を活用して発話タイミングの精度をより高めていく必要がある。

5. まとめ

本稿では、車内においてシステム主導型の音声対話を実現することを目的として、車載端末から受信した情報に基づいて発話を行う音声対話システムを実現するために必要となる、システム主導発話を行うための情報収集方式、及びシステム主導発話のタイミング制御方式について考察を述べた。今後の課題としては、システム主導発話を行うルール数が多くなった時の運用コストや、ユーザの負荷推定の高精度化が必要と考えている。

参考文献

- [1]. 藤本, 自然な発話により操作可能なカーナビゲーションシステムの開発, 電子情報通信学会論文誌, Vol.J96-D, No.11, 2815-2824(2013)
- [2]. 河原, 対話を通じてユーザーの意図・興味を探り情報検索・提示する情報コンシェルジュ, 情報処理, Vol.49, No.8, 912-918(Aug.2008)