

車載用音声対話インタフェースにおける対話戦略の評価

清信 将範 渡辺 隼平 川端 豪

関西学院大学理工学部 〒669-1337 兵庫県三田市学園 2-1

E-mail: {kiyonobu, amv82789}@ksc.kwansei.ac.jp, kaw@kwansei.ac.jp

あらまし 車載情報機器の操作性は走行の安全性に直結するため、そのインタフェースの良否を適切に評価することが重要である。本報告では、ドライビングシミュレータの運転中に音声対話によって機器操作を行う状況を想定し、いくつかの異なる対話戦略が被験者にどのような認知的負荷を与えるか検討する。今回はその端緒として、対話の進行の主導権の制御に注目し、NASA-TLXによる主観評価を試みた。現在までに、音声認識が完全な場合にはユーザ主導や混合主導の対話進行が好まれるが、認識率が悪くなるとシステム主導の対話進行が好まれるという妥当な結果を観察している。

キーワード テレマティクス、音声対話システム、NASA-TLX

Evaluation of dialog strategies for spoken-dialog-based telematics systems

Masanori KIYONOBU Shunpei WATANABE and Takeshi KAWABATA

Department of science and engineering, Kwansei Gakuin University 2-1 Gakuen, Sanda-city, Hyogo, 669-1535 Japan

E-mail: {cwz69524, amv82789}@ksc.kwansei.ac.jp, kaw@kwansei.ac.jp

Abstract Evaluation of telematics interfaces is important because it has a direct effect to safety driving. This paper describes an approach to measure the cognitive workload of spoken dialog strategies on simulation driving. Three types of dialog strategies (system, user, and mixed initiatives) are evaluated by utilizing NASA-TLX criteria. Generally, users may prefer to the user-initiative or mixed-initiative controls only when the recognition accuracy is complete.

Keyword Telematics, Spoken dialog system, NASA-TLX

1. はじめに

車載情報機器における音声対話インタフェースは、画面を注視せず（アイズフリー）、手をステアリングから離さずに使用できる（ハンズフリー）という利点から、運転中でも安全に利用できることを期待されている。しかし音声対話による機器操作においても、音声対話の制御法によっては思考が拘束され、安全が損なわれる場合がある。これはインタフェース設計の良し悪しが大きな要因であるため、対話戦略を適切に立てることが重要である。また、どのような音声対話が安全であり、逆にどのような音声対話が運転に悪影響を与えるのかを評価することは重要となってくる。

異なる対話戦略が運転者にどのような精神的負担を与えているかを評価した研究は過去にもいくつか行われている。清水ら[1]は対話形式の違い、西本ら[2]は対話場面の違い、宗近ら[3]は思考拘束形態の違い、野田ら[4]は提示する情報量の違いや目的意識の違い

について、それぞれ着目して評価を行っている。

本報告では、対話戦略を検討する手がかりとして、戦略をいくつかの要素に分解して考えることにする。まず、タスクがいくつかのスロットを埋めることによって進行される形式（Slot Filling Task）であった場合、そのスロットを埋める順序の主導権をどのように制御するかという観点がある。これを対話進行の主導権と捉えることにする。システムがスロットを埋める順序を1つ1つ指定するシステム主導型、ユーザが自由な順序でスロットを埋めることができるユーザ主導型、そしてそれらを併用する混合主導型などの方式が考えられる。また、音声認識には誤りがつきものなので、誤った場合にどのように修正を行うかについての対話戦略の要素もあり得る。さらに、受理した内容をシステムが確認発話するタイミングにもいろいろな方式があり得る。このように、対話戦略はいくつかの部分戦略要素に分解できる。

今回はこの第一の「対話進行の主導権」に注目し、

その評価を行う。このとき、車内環境雑音のため生じる認識精度の変化に注目し、適切な対話戦略はどう変化するかを調べる。

本稿では、2章で実験方法について、3章で実験の結果について、4章でまとめと今後の課題について述べる。

2. 実験方法

2.1. 評価対象

対話戦略を構成する要素には、対話進行の主導権、発話の自由度、応答のタイミング、確認の方法など、様々なものが考えられる。これらの要素またはその組合せが、車載用音声対話インタフェースの使いやすさにどのように寄与するかを調べるのが、本研究の目的である。今回はその端緒として、これらの要素のうち、対話進行の主導権に注目して評価を行う。対話進行の主導権には、システム主導、ユーザ主導、混合主導がある。それぞれの対話例を図1に示す。

どの対話戦略が適切であるかは、音声認識率の高低によって変動することが予想される。システム主導は煩わしく効率性に欠けるが、認識率が悪くても確実に対話が進められる。一方ユーザ主導は、認識率が高ければ非常に効率的であるが、誤認識が生じる場合には、認識結果の確認、訂正に際するユーザの精神的負担が増大することが予想される。そこで本報告では、対話進行の主導権と認識率との関係を調べるために、3つの主導権に対し、それぞれ音声認識率を変化させた音声対話インタフェースを被験者に使用させる。その際に被験者にかかる精神的負担を、NASA-TLX法を用いて主観的評価で測定する。NASA-TLXについては、2.4節で詳細を述べる。

2.2. タスク設定

対話進行の主導権と認識率との関係を調べるために、以下のようなタスク設定を行った。

【タスク】

被験者は、ドライビングシミュレータ（詳細は2.3節）を運転しながら、システムとの音声対話によるレストラン検索を行う。「地域」、「ジャンル」、「値段」の3つのスロットを埋め、システムに検

検索条件: 大阪, 和食, 2000円

○システム主導

Sys : 検索する地域を指定してください。

Usr : 大阪

Sys : 検索する地域は大阪府。よろしいですか？

Usr : はい。

Sys : 店のジャンルを指定してください。

Usr : 和食

Sys : 店のジャンルは和食料理でよろしいですか？

Usr : はい。

Sys : 予算を指定してください。

Usr : 2000円。

Sys : 予算は1000円でよろしいですか？

Usr : いいえ。

Sys : 予算を指定してください。

Usr : 2000円。

Sys : 予算は2000円でよろしいですか？

Usr : はい。

Sys : 大阪府の和食料理で、予算は2000円でよろしいですか？

Usr : はい。

Sys : 検索します。

○ユーザ主導

Sys : レストラン案内システムです。条件をどうぞ。

Usr : 大阪府の和食料理で、予算は2000円のところでお願い。

Sys : 大阪府の和食料理で、予算は1000円でよろしいですか？

Usr : いいえ。

Sys : もう一度ははじめからお願いします。条件をどうぞ。

Usr : 大阪府の和食料理で、予算は2000円。

Sys : 大阪府の和食料理で、予算は2000円でよろしいですか？

Usr : はい。

Sys : 検索します。

○混合主導

Sys : レストラン案内システムです。条件をどうぞ。

Usr : 大阪府の和食料理をお願いします。

Sys : 予算を指定してください。

Usr : 予算は2000円で。

Sys : 大阪府の和食料理で、予算は1000円でよろしいですか？

Usr : いいえ。

Sys : もう一度ははじめから。条件をどうぞ。

Usr : 大阪府の和食料理で、予算は2000円。

Sys : 大阪府の和食料理で、予算は2000円でよろしいですか？

Usr : はい。

Sys : 検索します。

図1. 対話進行の主導権の違いによる対話例

索を開始させた時点でタスクを終了とする。

【訂正方法】

システムの確認発話に対して、被験者は「はい」または「いいえ」のみで答える。

【確認方法】

3つのスロットが全て埋まった時点で、システムはユーザの入力発話に対して一括して確認を行う。ただし、システム主導の場合は、対話進行の性質上、項目ごとに毎回確認を行う。

【認識率】

認識率は、100%、80%、60%の3種類をWOZ法により与える。「はい」または「いいえ」による訂正発話に関しては誤認識を起こさず、3つのスロットへの入力に関する音声認識に対してのみ誤認識を発生させる。

図1に示す対話進行の主導権の違いによる対話例の中に、意図的に誤認識を混入させている。システム主導の場合は、システムが1つ1つ埋めるべきスロットを指定することによって対話を進める。ユーザ主導の場合は、ユーザが自由な順番で発話し、3つのスロットを埋める。ただし、全てのスロットが埋まるまでは、システムは入力待ち状態であり、その間システム発話はない。混合主導の場合は、初めはユーザに主導権があるが、3つのスロットが埋まらない場合はシステムが主導権を取り、足りないスロットについて質問をする。

最後の一括確認において誤認識が発生した場合には、もう一度初めから対話をやり直す設定にしている。これは、純粋に対話進行の効果のみを比較するためである。また、検索する条件(例:大阪, 和食, 2000円)は、1対話ごとに指定している。指定した内容を被験者が覚えたり思い出したりする負担をなくすために、走行中も被験者の視界の範囲内に検索条件を表示している。

2.3. ドライビングシミュレータの概要

運転状況を実験室内で容易に再現するために、パーソナル・コンピュータ上で動く簡易なドライビングシミュレータを作成した。ドライビングシミュレータの概観を図2に示す。操作デバイスは、ロジクール社製のステアリングコントローラー「GT FORCE Pro」を使用した。ハンドルの半径は25cm, 回転角は360度に設

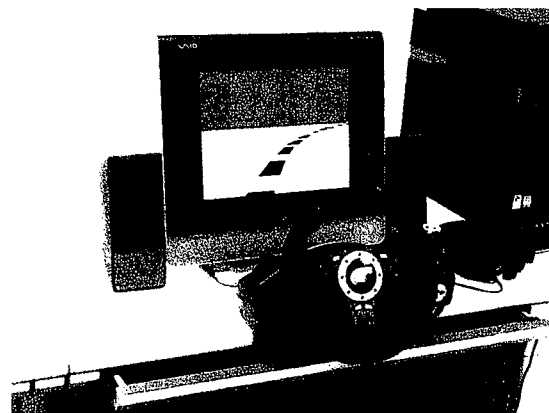


図2. ドライビングシミュレータの概観

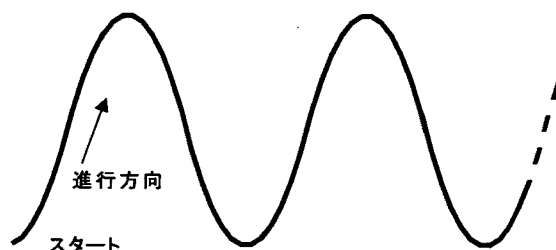


図3. 走行コースの俯瞰図

定した。ディスプレイのサイズは17インチで、被験者との距離は60cm程度とした。走行コースの俯瞰図を図3に示す。コースの最大曲率は半径50mである。直線から左右の最大曲率まで一定の割合で曲率を変化させ、被験者がハンドルの角度を常に調節しなければならない状況を設定した。車両の速度は時速50km/hで固定している。すなわち、被験者はハンドル制御のみを行い、アクセル制御は行わない。

定常運転課題では、被験者の慣れにより、運転に費やす注意量が減少していく[5]。しかし実際の運転では周辺状況の変化に対応する注意量が必要であるため、実験室内においても運転に必要な注意量を確保する必要がある。そこで本研究では悪路走行を想定し、走行中にランダムな妨害(強制的車両向き変更)を発生させ、被験者が常に状況変化に対応しなければならない状況を作った。どの程度の妨害頻度・強度が適切であるかは、前報[6]において、3秒に1回、左右に0度~8度の間でランダムに発生させることを導出している。今回の実験においても、この妨害頻度・強度

を用いている。

2.4. NASA-TLX

NASA-TLX (NASA Task Load Index) とは、作業に伴う精神的負担を、被験者本人により主観的に評価する方法である[7]。NASA-TLX は、知覚的要求、身体的要求、タイムプレッシャー、作業成績、努力、フラストレーションの6つの尺度から構成されている。評価の手順は以下の通りである。

- (1) 負担度評価：被験者は6つの尺度それぞれに対して「高い/低い」の両極間に引かれた線分上の適当な位置にチェックを入れる。実験者はその位置を0～100の数値に変換し、各尺度の評定値とする。
- (2) 重要度判定：被験者は6つの尺度全ての組合せ(15対)に対し、作業負担の要因としてより重要だ

と思う方を選択する(一対比較)。一対比較の中で選ばれた回数(最大5, 最小0)を、各尺度の重要度とする。

(3) WWL 得点の算出：

(1) で得られた評定値を、(2) で得られた重要度を用いて加重平均した値(WWL 得点：mean weighted workload score)を NASA-TLX の総合値とする。

NASA-TLX において精度の高い作業負荷の評価を行うためには、被験者が6つの評価尺度を十分に理解できるように、評価尺度の説明を実験に合わせて簡易化及び具体化する必要がある [4][7][8]。

本研究でも従来研究を参考に、評価尺度の説明文や負担度評価の際に表示する終端点の説明を、実験のタスク内容に合わせて簡潔かつ具体的にし、被験者の困惑や誤解のないよう心がけた。表1に各尺度の説明文の内容を示す。

表1. NASA-TLX における尺度名とその説明文

尺度名	説明文(上段は負担度評価, 下段は重要度判定に使用)	終端点の説明 (上段は最小, 下段は最大)
知覚的要求	聞き取ったり、判断したりするのは大変でしたか？	簡単だった
	どの程度の知覚的活動(判断する、聞き取る、など)を必要とするか。課題が易しいか難しいか。単純か複雑か。正確さが求められるか大ざっぱでよいか。	大変だった
身体的要求	しゃべること(声を出すこと)は大変でしたか？	簡単だった
	どの程度の身体的活動(しゃべること、声を出すこと、など)を必要とするか。休み休みできるか働き詰めか。	大変だった
タイムプレッシャー	時間的切迫感ほどの程度でしたか？	余裕があった、ゆっくりできた
	時間的切迫感がどの程度か。ペースはゆっくりとして余裕があるものか、それとも焦って余裕のないものか。	余裕がなかった、焦った
作業成績	うまくやり遂げられましたか？	達成できた、満足している
	課題の目標をどの程度達成できたか。目標の達成に関して自分の作業成績にどの程度満足しているか。	十分は達成できず、満足していない
努力	一生懸命になりましたか？	努力するほどではなかった
	与えられた課題の維持・達成に、どの程度一生懸命にならなければならないか。	かなり努力した
フラストレーション	イライラ・ストレス・不安感などをどの程度感じましたか？	リラックスしていた
	イライラ、ストレス、不安感、落胆、悩みをどの程度感じるか。あるいは逆に安心感、満足感、充足感、楽しさ、リラックスをどの程度感じるか。	ストレスを感じた

また、被験者の判断を支援し、負担度評価における数値変換を詳細かつ正確に行えるよう、芳賀ら[7]を参考に、GUIによるコンピュータソフトを作成した。図4は負担度評価の画面の一例である。被験者は画面中央のスクロールバーを左右に動かし、適当な位置で確定する。そのときの座標が0~100の評定値としてファイルに書き出される。図5は重要度判定の画面の一例である。被験者は画面下部に書かれた各尺度の説明を確認しながら選択することができる。

2.5. 実験手順

被験者は20代の男性2名（ともに運転歴5年）である。3種類の対話進行の主導権と3段階の認識率との組合せで、9通りの音声対話をそれぞれ5回ずつ行った。実験の手順は以下の通りである（図6）。

(1) ドライビングシミュレータの運転練習：
1分半の走行を10回程度行う。

(2) 運転しながら音声対話タスクを行う

(3) NASA-TLXの負担度評価

(4) NASA-TLXの重要度判定

※ (2) ~ (4) を9通りの音声対話について行う。

(5) 15分間の休憩

(6) ドライビングシミュレータの運転練習を1回。

※ (2) ~ (6) を5回繰り返す。

実験終了

9通りの音声対話はランダム順に行った。

3. 実験結果

図7は、9通りの音声対話における、各被験者のNASA-TLX総合値を示している。縦軸方向に大きいほど、その対話における精神的負担が大きいことを表す。棒グラフに付加しているヒゲは、5回分のデータの標準偏差を表している。被験者間でNASA-TLX総合値に差があるのは、NASA-TLXが主観的に評定値を与えるものであるため、被験者ごとに数値の基準が異なるからである。しかし両者の傾向はほぼ一致している。

認識率が100%の場合はユーザ主導や混合主導の方が精神的負担が低い、誤認識が混入すると逆転してしまっている。これは、100%のときはユーザ主導や混合主導は非常に効率がよく、短時間でタスクを達成することができるのに対し、システム主導は1つずつ

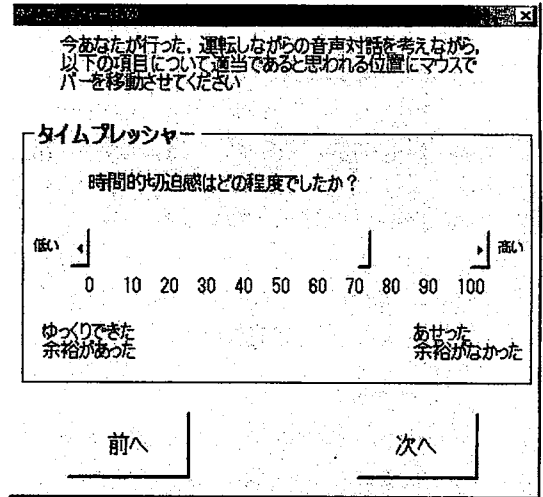


図4. 負担度評価の画面

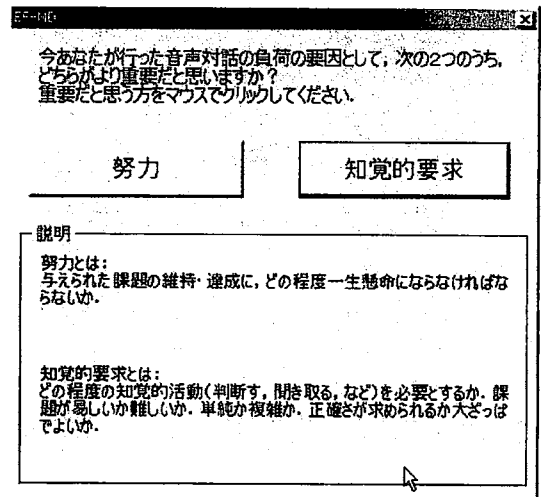


図5. 重要度判定の画面

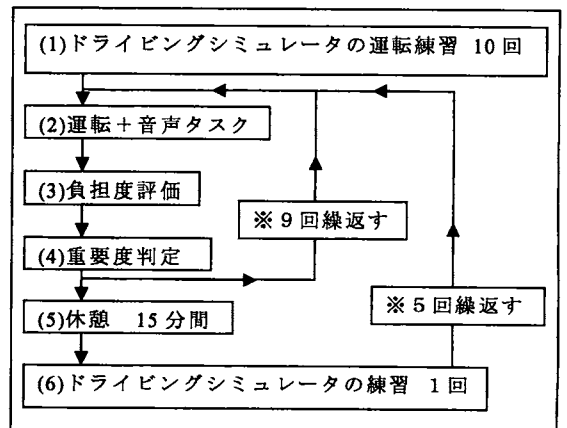


図6. 実験の流れ

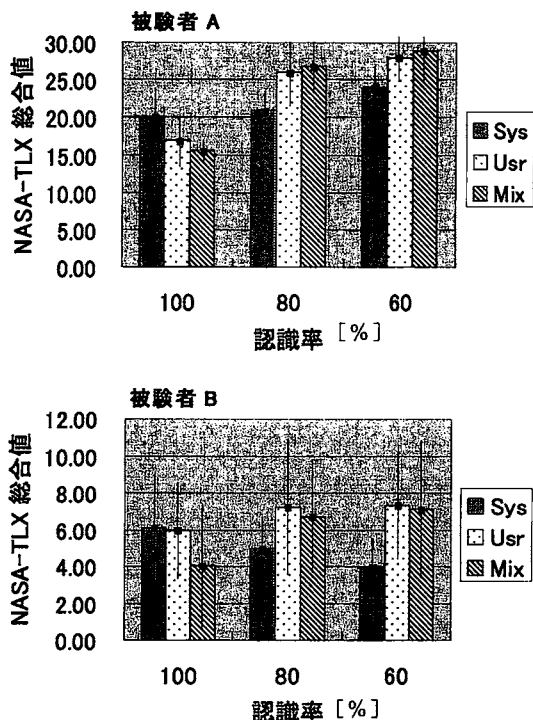


図7. 被験者ごとのNASA-TLX 総合値

ロットを埋めていくことに煩わしさを感じるからだと考えられる。しかし誤認識が起きると、ユーザ主導や混合主導ではもう一度初めから言い直すことになるため、精神的負荷が大きく増大するのに対し、システム主導では1スロットずつ確実に入力が行えるため、誤認識が起きても精神的負荷が大きく変化することはない。その結果、このような逆転が起きていると考えられる。ユーザ主導と混合主導にあまり差がないのは、混合主導において3つのスロットを一度に入力する事例がほとんどで、対話内容がユーザ主導とほぼ一致したからである。

これらの結果から、認識率が高ければユーザ主導が好まれ、認識率が悪ければシステム主導が好まれるだろうと予想したことが、具体的な実験によって確認された。

しかし両方の被験者において、各音声対話内でのばらつきが小さいとはいえない。これは両被験者が、実験を重ねるにしたがって、負担度の評価を低く評定していく傾向が見られたからである。この順序効果の影響を除去した分析を今後行う必要がある。

4. おわりに

本研究では、車載用音声対話インタフェースにおいて、ドライバーの精神的負担の少ない音声対話を実現することを目的とし、今回はその端緒として、対話進行の主導権の制御に注目し、対話戦略の評価を行った。その際、車内環境では避けられない誤認識の影響も考慮し、認識率の変化によって適切な対話戦略がどう変化するかを調べた。実験ではドライビングシミュレータを用いて運転状況を再現し、これを運転させながら音声対話タスクを行わせ、NASA-TLXによる主観評価で精神的負担を測定した。その結果、認識率が高い状況ではユーザ主導が好まれ、認識率が悪い状況ではシステム主導が好まれるという妥当な結果を、具体的な実験により確認することができた。

今後は、今回行った実験を更に被験者を増やし、結果の信頼性を高めていく予定である。更に、確認の方法、訂正方法、応答のタイミングなどの、対話戦略を構成する他の要素についても検討を進めていく。

文 献

- [1] 清水司, 小島真一, 脇田敏祐, 本郷武朗, “運転中における音声対話システムの評価,” 情報処理学会研究報告, 2000-SLP-32-17, pp.87-92, 2000.
- [2] 西本卓也, 高山元希, 櫻井晴章, 荒木雅弘, “音声インタフェースのための対話負荷測定法,” 2004/02 電子情報通信学会論文誌, vol. J87-D-II No.2.
- [3] 宗近純一, 松坂要佐, 小林哲則, “車運転時における音声利用の心的作業負荷,” 第2回情報科学技術フォーラム FIT2003 情報技術レターズ, vol.2, pp.105-106 Sep, 2003.
- [4] 野田幸志, 西山昌史, 堀内靖雄, 市川薫, “信的負荷状況における車載情報機器のための音声対話戦略の分析,” 情報処理学会研究報告, 2006-SLP-64, pp.149-154, 2006.
- [5] 海保博之, 加藤隆, “認知研究の技法,” 福村出版, 1999.
- [6] 清信将範, 川端豪, “ドライビングシミュレータの走行精度による車載情報機器の負荷測定,” 信学技報, SP2006-13, pp13-18, 2006.
- [7] 芳賀繁, “メンタルワークロードの理論と測定,” 日本出版サービス, 2001.
- [8] 萩原亨, 徳永ロベルト, “メンタルワークロード評価法に基づく運転中の携帯電話利用の影響に関する研究,” 国際交通安全学会誌, vol.30, No.3, pp.66-73, Sep 2005.