

ナビ対話におけるユーザ応答時間と満足度*

☆網本侑理（上智大），神沼充伸（日産自動車），荒井隆行（上智大）

1 はじめに

現在普及している車載用音声インターフェイス(以下 I/F)は、ユーザがシステム内で設定された語（音声コマンド）と完全に一致する言葉を発話しないとシステムが動作しない。このような I/F をユーザが使いこなすには、予め音声コマンドを覚えた上、地名などの固有名詞を正式名称で、かつ、フィルター抜きに入力しなければならない。こうした不便さを解決する方法として、自然言語理解(NLU : Natural Language Understanding)や対話管理(SDM : Spoken Dialogue Manager)が提案されており¹⁾、これら技術の導入によりユーザは入力上の制約を感じにくくなることが期待される。既存の設計方法としては、設計者がユーザとシステムの対話状況を想定して設計する手法や POMDP のような機械学習を用いた手法が提案されている。一方、車載用では、ユーザが使用する機能を最後まで確実に達成する要件を満し、かつ、ユーザビリティの観点から十分なユーザ満足度が得られている必要がある。然るに、インターネット上の音声入力サービスはこれら要件を満たしておらずそのまま車載することは困難である。そこで本稿では、先ず NLU/SDM を導入した車載音声 I/F のデザインおよび評価手法について改善すべき点を検討し仮説を創出した。次に、ユーザの満足度を向上させるような対話デザインを提案し WOZ²⁾を用いたシステムで提案手法の効果を評価した。

2 音声 I/F 改善点の検討

2.1 デザイン上の改善点の抽出

先行研究で収録された、運転中のナビ操作課題（WOZ²⁾、エアコン操作等、29名分）の対話を筆者が Praat を用いて定性的に分析したところ以下の可能性が示唆された。

1) システムガイド音声終了後、ユーザが反応するまでの時間が長いと不満がある。

2) 発話入力後、システム応答のテンポが悪い（応答が遅い、長すぎるなど）とユーザ応答し難くなりシステムに対する印象が低下する。

2.2 デザインの良さを測る評価尺度の検討

新たなデザインの正当性を示すには、適切に評価できる評価尺度を創出する必要がある。従来は、システムの受諾性能を測る認識率（WER : Word Error Rate 等）、ユーザビリティの良さを測るタスク達成時間（TTT : Total Task Time）が知られているが、2.1 節を鑑み、ユーザ応答時間と満足度の関係を細かく調べるため以下の評価尺度を創出した。

- 1) 満足度を直接聞く質問紙による主観値
- 2) システムガイドの分かりやすさ、システムのテンポに寄与する指標（Latency, ユーザ応答の容易さ）
- 3) 分かりやすさの指標としての入力エラー通知(3.1 節の Explain 発話)後のユーザ発話応答時間
- 4) 明瞭性の指標としての確認動作(3.1 節の Confirm 発話)後のユーザ発話応答時間

2.3 対話 I/F の設計指針に関する提案

2.1 節で述べたように、テンポ良く対話システムを使えるようにするためには、使い方が明確でユーザが応答しやすいことが望ましい。例えば、Levinson³⁾らは、話者交替時の応答時間は、話者の思考時間や交替キューの不明瞭さなどに影響されて長くなることを指摘している。ユーザがあまり思考せずに返答でき、発話準備のための負荷が少ない対話システムを作ることができれば、高い満足度に繋がる可能性がある。そこで我々は、ユーザの応答しやすさ＝ユーザ応答時間と定義し、「ユーザ応答時間が短くなるように対話をデザインすることでシステムに対するユーザの満足度が上がる」と仮説を立てた。3 節以降では、ユーザ発話応答時間を短くするための工夫をした対話システムを疑似的に構築し、ユーザ応答時間に注目し、対話システムを設計する

* Relationship between Findings from Users' Response Time and Satisfaction Level in Vehicle Dialogue System, by AMIMOTO, Yuri (Sophia Univ.) KAMINUMA, Atsunobu (Nissan Motors) and ARAI, Takayuki (Sophia Univ.).

ことが満足度の向上に繋がることを実験的に示す。

3 対話システムの構築と対話収集

3.1 対話システムの構築

車載用音声 I/F の応答を Table1 の 5 種類へ分類した。Confirm はユーザの発言内容の確認、Please はコマンド入力の催促、Explain は入力に応答できないときの説明、Question は Value を引き出すための質問をする。また Answer はユーザのコマンド入力を受け、システム動作や検索内容などの返答をする。Table1 では各発話内容を、既存の車載用音声 I/F を模した従来法と、ユーザ発話応答時間を短くする工夫をした制御法とで対比させてまとめた。

Table1 対話システムの行う発話

分類	発話内容
Confirm	(従来)「でよろしいですか」 (制御)「ですか」「ですね」
Please	(従来)「コマンドをどうぞ」 (制御)「何がしたいですか」
Explain	(従来)「検索の結果 XX は見つかりませんでした。登録されていないか入力の誤りである可能性があります。」 (制御)「見つかりませんでした」「わかりません」
Question	(従来)「室温を何度に変更しますか」 (制御)「何度にしますか」
Answer	(従来)「設定温度を 27 度に変更しました」 (制御)「はい」「27 度にしました」

以下の Fig.1 は制御法の対話システムで応答する WOZ 画面の例である。



Fig.1 制御法の WOZ システム

制御法は、システム発話を短く、コマンド催促を Wh-Question の形式で行い、ユーザの発言に相槌を打つ、といった構成にした。

3.2 対話データ収集の枠組み

16名(14M; 2F、20代から50代)のドライバーにドライビングシミュレータ (Fig.2) による疑似運転課題を課した環境下で、対話システムを使用することを求め、対話データを収集する。ドライバーは4つの課題(エアコンの温度設定、メール受送信、目的地設定、天気予報の検索)を、システムと対話しながら解決する。4つのうち2つは従来の対話システムの応答、2つは提案する対話システムの応答がなされ、課題およびシステムの組み合わせは無作為に設定された。このときドライバーにはシステム応答に違いがあることは伏せた。



Fig.2 ドライビングシミュレータ

ドライバーは始めにドライビングシミュレータに慣れるため数分間運転を行い、次にシステムとの対話に慣れるために音楽再生の練習課題を行う。本課題では、課題毎に状況(車内が暑くなってきたので温度を下げたいと思っている等)を説明することでドライバーに状況理解を促し、その後、走行を開始、課題解決を図る。走行中にも課題内容は確認できるよう、Fig.2の右下のディスプレイに要約を表示する。課題はドライバーが課題を達成するか、終了を求めるまでとする。課題終了毎に停車し、システムの主観評価を実施する。

3.3 主観評価手法

満足度に関する評価の質問項目を Table2 に纏める。各項目に対し、ドライバーには「非常に不満」「やや不満」「やや満足」「非常に満足」のいずれかを選択するように求めた。

Table2 直後質問

評価項目	質問
時間満足度	「達成できるまでの時間は満足できましたか」
認識精度	「このナビは自分の入力した」

	言葉を正しく認識できなかった」
明瞭性	「このナビの音声ガイドの内容が理解できた」
応答時間	「このナビの応答時間は適切だった」
分かりやすさ	「音声ガイドの後に何を入力するべきかすぐにわかった」

3.4 発話内容のラベリング

ユーザは特定のシーンにおける動作についてだけ性能差を感じている可能性がある。このような対話の流れおよび発話の内容毎の間を分析するため、収集したユーザ発話に内容ラベリングを行った (Table3)。

Table3 ユーザ発話内容ラベル

分類	発話例
Call	「ナビ」
Command	「音楽」「目的地検索」「暑い」
Value	「西丹沢の天気」「星野源」「27度」「駅周辺のコンビニ」
Yes	「はい」「そうです」「それで」
No	「いいえ」「だめ」「違います」
Filler	「えっと」「はいはい」「あっ」

Call は発話開始時にユーザがナビを呼ぶ発話、Yes/No はシステムからの確認に対するユーザ応答である。Command と Value はシステムに指示を出す発話で、固有名詞を含む場合には Value とした。Filler は、システムの応答以外の発話とした。システムの発話内容は 3.1 節の通り、5 種類に分類した。

4 結果の分析と考察

以下統計に学生 t 検定を用いた。

4.1 手法間の評価差

制御法はユーザが反射的に返答できることを目指した、相槌を打ち、端的な返答を返す対話デザインなので、ユーザには発話のタイミングや内容が分かりやすくなっている。提案した対話デザインが目的に沿うものであれば、評価項目の「分かりやすさ」「明瞭性」のスコアが従来法より改善されることが期待される。手法別にスコアの平均をみると (Fig.3)、制御法は従来法と比較して「分かりやすさ」「明瞭性」「時間満足度」で有意に高いスコアが出ており、満足度が向上していた。これらの結果から、制御法が「分かりやすさ」「明瞭性」「時間満足度」において効果があることが示唆された。一方、「応答時間」と「認識精度」

は、手法による有意なスコア差が検出されなかった。このことは、2つの項目は何れもオペレータの手腕に依存し、デザインとは無関係であることが要因と考えられる。

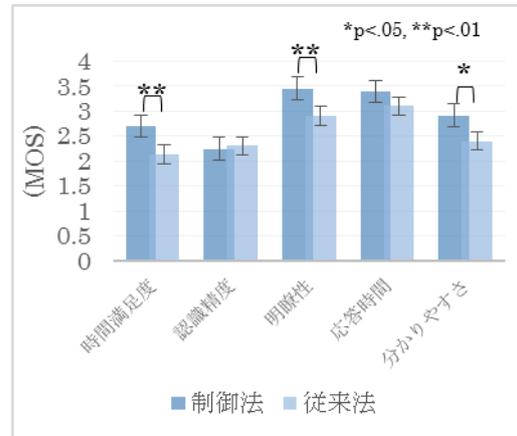


Fig. 3 従来法と制御法の評価スコア

4.2 Explain 発話に対する発話の分析

Explain 発話から、ユーザはエラー時のシステムの解釈、状態を知ることができる。分かりやすいシステムとの対話では、Explain 発話の後にも何を答えるべきかユーザがすぐに分かり、ユーザ応答時間の平均および分散が、小さくなると予想される。ここでは 2.2 節で評価尺度として挙げた Explain 発話後のユーザ発話応答時間を「分かりやすさ」を物理的に知る尺度と見て分析する。

制御法では、システムは短くシンプルな Explain 発話を行う。このときユーザはシステムの発話に影響されてシンプルで短い応答を返しやすくなり、その結果「分かりやすさ」の評価が高くなると考えている。Fig.4 に Explain 発話後のユーザ応答時間の平均および分散を、Fig.5 に主観評価の結果 (当該発話を含む課題の結果) を示す。Fig.4 では手法間で平均には有意な差は検出されなかったが、分散では制御法が有意に小さくなっていた ($p=.029$)。このことから、提案法の短い Explain 発話により、課題を通して解釈に掛かる時間が安定したと考えられる。一方、従来法では Explain 発話が長いいため同一課題内でも応答時間の分散が大きく、ユーザ解釈の時間が変動しやすい。また、平均に有意差が無かったのは、手法によらず課題の難易度やユーザの状況によって応答時間は左右されるため、Explain 発話時間を短くしただけでは複雑な状況を解決し得ず、有意な差が出なかったと

考えられる。Fig.5 の主観値は制御法において「分かりやすさ」のスコアが有意に高い ($p=.017$)。このことから Explain 発話後のユーザ応答時間の分散が、「分かりやすさ」の主観値に寄与する可能性が示唆された。

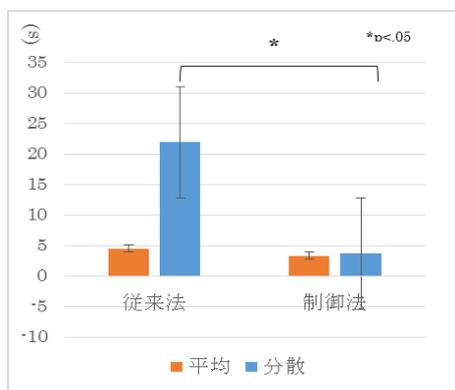


Fig.4 Explain 発話後のユーザ発話応答時間

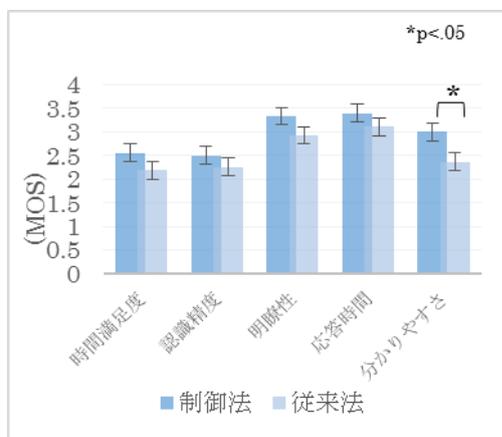


Fig.5 Explain 発話後の手法別評価スコア

4.3 Confirm 発話に対する発話の分析

制御法の Confirm 発話も短くシンプルな発話になるようにした。システムの確認内容がユーザにとって明瞭であれば、返答しやすくユーザ応答時間も安定すると考えられる。また、短い応答時間から課題時に掛かった時間も短く感じられて「時間満足度」のスコアが高くなる期待がある。Fig.6 に Confirm 発話後のユーザ応答時間の平均および分散を、Fig.7 に主観評価の結果を示す。Fig.6 では制御法と従来法を比較すると応答時間の平均が有意に小さくなった ($p=.014$)。このことから制御法は従来法と比べユーザにとって明瞭であったといえる。また、「明瞭性」と「時間満足度」が向上していることから (Fig.7)、Confirm 発話後の解釈にかかる時間はユーザ応答時間から推測でき、ユーザ解釈時間の長さは主観的な明瞭性に寄与する可能性が示唆された。

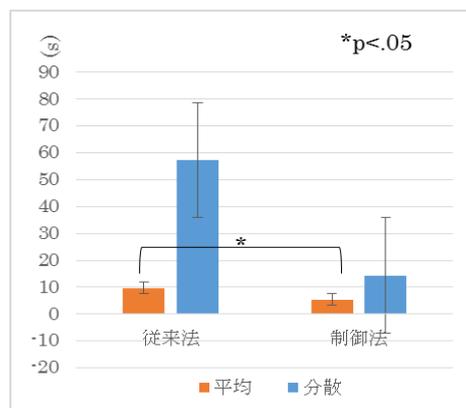


Fig.6 Confirm 発話後のユーザ発話応答時間

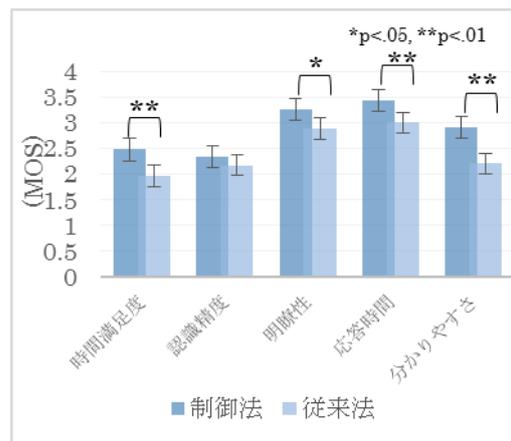


Fig.7 Confirm 発話後での手法別評価スコア

5 おわりに

本稿では対話 I/F の指針にユーザ発話応答時間を用いることについて検討した。Explain 発話後のユーザ応答時間の分散を抑えることが「分かりやすさ」、Confirm 発話後のユーザ応答時間の平均を小さくすることが「明瞭性」の向上に繋がっている可能性が示唆された。制御法ではユーザ応答時間を間接的に制御できたため主観的なスコアが向上した。

謝辞

2.1 節の対話データを提供して下さった同志社女子大学の和氣教授に感謝します。

参考文献

- [1] 神沼, 清水, 自動車技術論文集, Vol47,3,2016.
- [2] Bensen, Dybkjær, et al., "Wizard of oz prototyping: How and when." *Proc. CCI Working Papers Cognit. Sci./HCI*, 1994.
- [3] Levinson, *Trends in cognitive sciences* 20.1, 6-14, 2016.