

ドライバの負担を考慮した車載音声インタフェースの対話デザイン検討

Study of spoken dialogue for speech interface in vehicle considering driver's workload

清水 俊行
Toshiyuki Shimizu

神沼 充伸
Atsunobu Kaminuma

日産自動車株式会社
Nissan Motor co., Ltd.

Speech interfaces based on natural language understanding and spoken dialogue management are continually becoming more sophisticated functionally, and their implementation in vehicles is being studied. In envisioning the use of speech interfaces in vehicles, it will be necessary to factor the associated driver workload and usability during driving into the fundamental design. Toward that end, we undertook a study of driver workload when using a speech interface, which is a subject that has not been sufficiently discussed so far. Various issues involving speech interfaces having different specifications were reproduced on a driving simulator. An evaluation of the influence on the driver's physiological aspects revealed that inputting voice commands in chunks helped to mitigate higher-order brain function. This finding is assumed to be fundamental knowledge concerning the design of conversational interfaces for vehicle application. It can serve as one validation method for the implementation of speech interfaces in vehicles.

1. はじめに

スマートフォンやロボットなどに対して、自然言語理解や対話管理機能を使った音声インタフェース(以下、IF)の組み込みが進んでいる。音声 IF の機能は、正しく認識ができること(認識率)、正しい応答がなされること(受諾率、応答正解率)などが評価され、機能の改善を目的とした知能化の研究が進められている [Hilton 2012], [Sagae 2009].

一方、雑音環境下である車の HMI (Human machine Interface) においても自然言語理解や対話管理機能を使った音声 IF の導入が検討されるようになってきた。音声 IF の車載化のためには、従来、実施されてきた認識率などのシステムの機能評価だけでなく、一定時間内に入力が完了すること(達成時間)や入力が達成でること(達成率)など Usability に関する要件も求められる。更に、音声 IF の使用がドライバの負担にならないこと、応答タイミングなど、安全性に関する機能の評価も考慮しなければならない。

先行研究では、神沼らが、音声 IF がもつ課題を明示し、音声入力課題の達成時間、達成率、および質問紙法による心理尺度の分析を用い、自然言語理解や対話管理機能を用いた音声 IF を運転中に使用することの有効性を議論している [神沼 2015 春・秋]。しかしながら、神沼らの研究はドライバ負担の観点からの検討が十分に行われておらず、運転中に適用可能であることを示すには十分ではない。そこで、筆者らはこれまで車載用音声 IF の評価では十分に議論されていなかった、自然言語理解や対話管理機能を用いた音声 IF のドライバ負担について、定量的な評価を行い、実用性について考察する。

2. 車載用音声インタフェース

2.1 コマンド入力方式

現在普及している車載用音声 IF は、目的地設定、車載機器の操作、電話、楽曲検索等の設定に際し、特定の発話に対して

一つの機能(Transaction)が結びつけられている。例えば「目的地」と発話した場合は、ナビゲーションシステムの目的地入力ボタンを押した場合と同じ結果が得られる。ただし、ユーザは、機能と結びつけられている言葉(以下、コマンド)、地名などの読みおよび入力順序を正確に、かつ淀みない発話で入力することを求められ、使用前にコマンドを覚える必要がある。また助詞、動詞、接尾辞の違いやフィラー発話の存在は、誤認識の要因となり車載での利便性を低下させる。

2.2 自然言語理解と対話管理機能

コマンド入力方式の課題に対する改善手段として、自然言語理解(NLU: Natural Language Understanding)と対話管理(SDM: Spoken Dialogue Management)の技術に着目する。自然言語理解は、一部の欠けた情報やナビゲーションシステムで想定される語順と異なる文でも意味や意図を抽出する。この機能を想定すると、「箱根資料館へ行く」という発話入力については、「箱根資料館」を「箱根町立郷土資料館」、「～に行く」を「目的地設定」と読み替えて話者の意図を正確に受諾することを可能にする。また、対話管理機能は、不足情報があつた場合、システム側から問いかけ、追加の情報収集を行うことができる。例えば、

「目的地設定 箱根資料館」と発話したときに、何らかの理由で「目的地設定 箱根」と受諾してしまった場合、「施設名をもう一度お願いします」と聞き返すことでユーザのさらなる入力を促すことができる。これら2つの機能性を車載用音声 IF に導入することで、ドライバ負担を低減させる可能性がある。

そこで本稿では、ドライビングシミュレータ上に自然言語理解と対話管理機能を想定した音声 IF の受諾応答システムを再現し、音声 IF 使用時のドライバ負担、特に生理面への影響を調べる。

3. 評価実験

3.1 実験の構成

ドライビングシミュレータ(以下、DS)上で運転操作を行いながら、音声 IF の機能を操作する課題を実施した。また運転への集中度合いを調べるため運転席前方に取り付けた LED 点滅に

連絡先: 清水俊行, 日産自動車株式会社, 〒243-0123
神奈川県厚木市森の里青山 1-1,
Email: shimizu-toshiyuki@mail.nissan.co.jp

よる光刺激に対する反応課題も同時実施した。課題実施中は、ドライバの発話及び、ナビゲーションの応答を音声として記録し、光刺激への反応時間、ドライバの心電図、皮膚電気活動、脳血流の計測を行った。図1に実験のシステム構成を示す。

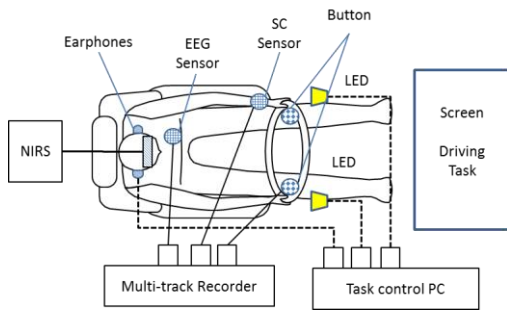


Fig1. Layout of experiment device

3.2 音声課題

音声 IF の比較は、ナビゲーション(以下、ナビ)操作で行い、音声課題としてナビで使用頻度の高い目的地検索(レストランなどの検索課題)と電話帳検索(目的とする相手へ電話を掛ける課題)を設定した。音声 IF は下記に示す(1)~(3)の異なる仕様を用意した。音声応答のルールについて説明する。

(1)Command Input

正確な目的地の名称が予め決められた順に発話することで入力できる。

(2)NLU Input

自然言語理解を用い、正確な名称に加え、人間が容易に解釈できる程度に曖昧な指示発話、通称および部分的に一致した名称を、正しい入力順または 倒置された入力順に発話することで入力できる。

(3)NLU/SDM Input

自然言語理解と対話管理機能を用い、NLU Input の機能に加え、ドライバの意図を推定し、確認のための応答を呈示する。



Fig2. Sample of Wizard of Oz tool

音声の応答は、ナビの音声対話機能を WOZ(Wizard of Oz)方式[Fraser 1991], [Berssen 1994]で模擬的に再現する。WOZ ツールは、画面上に読み上げる文章が書かれた複数の

ボタンが配置され、オペレータが押したボタンに記載された文が音声ガイドとしてドライバのイヤホンに呈示される。ボタンに記載された文は、Fig.2 のように完結した文章になっているものと、固有名詞、助詞、動詞が分割して配置されているものがある。読上げ可能な文および画面は課題毎に作成した。

オペレータは、音声 IF 仕様のルールに従い、ドライバの発話に対して WOZ ツールで適切な応答できるように十分な練習を行った。

3.3 実験の手順

実験は日産自動車(株)実験倫理審査委員会による承認のもと実施した。ドライバ役には事前に文章による実験手順を詳細に説明し、実験参加に関するインフォームドコンセントを得て実施した。ドライバ役は、普段から車の運転を行っている 30 名(男性 29 名, 女性 1 名, 20-50 代)である。

運転は、時速 80 km を目安に無理のない走行をしてもらうように教示した。十分な運転練習を行い DS の操作にも慣れてもらった。最初に 60 秒間の運転走行と光刺激反応課題のみで記録を行い、次に 3 つの音声 IF の仕様ごとに音声課題を 3 回(順に、目的地設定、目的地設定、電話番号検索)を行い、計 9 課題を実施した。音声 IF の仕様は、ドライバごとに順序入れ替えた。課題は入力が達成されればそこで終了し、達成されなくてもドライバ負担軽減のため1課題あたり、120 秒で打ち切った。

4. 結果

4.1 自律神経系への影響

心電図と皮膚電気活動は 11.025kHz で発話音声と同時にマルチトラックレコーダーへ取り込んだ。心電図は、胸部単極誘導にて計測を行い、心臓が収縮するときの主要な波である R 波の検出を行い、R 波と次の R 波の間隔(RR 間隔)を求めた。RR 間隔は 0.5 秒で再サンプリングを行い、周波数変換を行った後、LF 成分(0.04-0.15Hz)と HF 成分(0.15-0.45Hz)に分け、HF と LF/HF の平均値を求めた[大須賀 1993]。皮膚電気活動は、左手中指と人差し指から通電法により計測を行った[西村 1984]。皮膚電気活動は 0.04Hz でローパスフィルタを掛け、皮膚コンダクタンス水準 SCL を求めた。それぞれの指標は、運転のみの課題では、運転が安定した 60 秒間のデータを解析対象とし、発話をともなう課題では、課題提示後、発話開始から課題完了まで、もしくは課題が完了せず途中で打ち切った 120 秒までのデータを解析の対象とし、すべての課題で平均値を特徴量とした。また個人差による影響を排除するため特徴量は標準得点化(Z-value)で正規化した。

$$Z\text{-value} = (\text{実測値} - \text{平均値}) / \text{標準偏差} \quad (1)$$

初めに、運転のみ(W/O Input)と音声 IF 使用時の比較を行う。HF, LF/HF の結果を図 3 と 4 に示す。一要因分散分析の結果、HF と LF/HF に 1% 以下で有意差が認められたため、Tukey-HSD 法による多重比較を行った。(以降の下位検定はすべて、Tukey-HSD 法を適用する。)運転のみと比べ音声 IF 課題の実施により HF が減少し、LF/HF が増加を示した。SCL には有意な変化は見られなかった。

HF は副交感神経活動を反映する指標であり、リラックスしている時に増加を示す。HF の減少は音声 IF 使用による負荷が反映したと考えられる。LF/HF は交感神経活動の指標あり、緊張している時などに増加を示すが、呼吸活動の影響を受け、呼吸が不規則になることでも増加する。SCL も交感神経の活動を

反映し、覚醒水準の指標としても使われている。ここで SCL に有意差がないことから強い交感神経の促進がないとすると、LF/HF の増加は、発話による呼吸の乱れが強く影響したと考えられる。すなわち音声 IF の操作は、運転のみ(W/O Input)と比べ副交感神経の抑制が見られるが、交感神経を活性化させるほどの作業負荷は受けていないと考えられる。

検証のため、3つの音声 IF の仕様 Command Input, NLU Input, NLU/SDM Input 間だけで比較を行ったが HF, LF/HF 指標に差異は見られなかった。

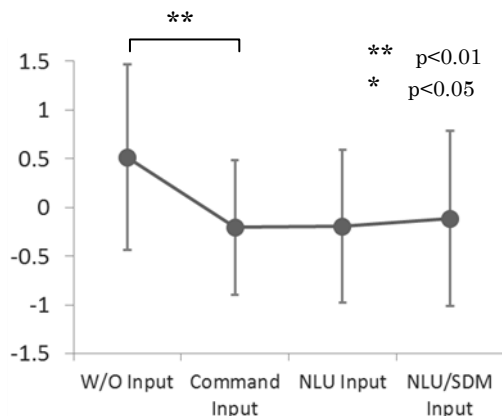


Fig3. Z-value of HF
(ANOVA, $p=0.0052$, $F=2.71$)

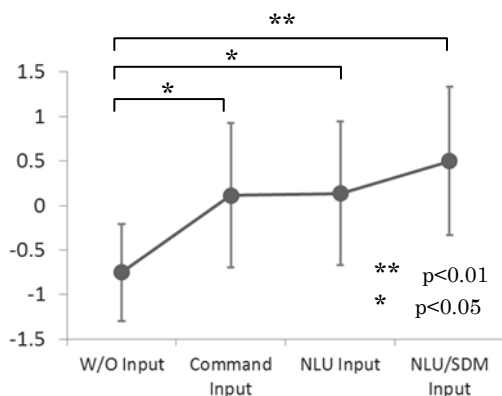


Fig4. Z-value of LF/HF
(ANOVA, $p=0.0002$, $F=7.86$)

4.2 脳活動の変化

次に、脳機能活動の観点から3つの音声 IF 間の比較を行った。脳活動の計測には時間的な制約があり、実験参加者のうち計測の合意が得られた 12 名のみで計測を行い、正常に計測できたデータ 10 名分で解析を行った。計測には、体動の制限が少なく自然な状態で実験が行える機能的近赤外分光法装置(島津製作所製 FOIRE-3000, 以下 fNIRS)[小西 2000], [網田 2005]を用いた。fNIRS 装置は、頭表から 2cm 程度の脳表組織の脳内ヘモグロビン(Hb)の変化が観測できる。本稿では図 5 のように前頭額部の 32 箇所からサンプリング間隔 0.175 秒で脳内 Hb の計測を行った。図中の数字位置は脳内 Hb を検出する位置のチャンネル番号を表す。

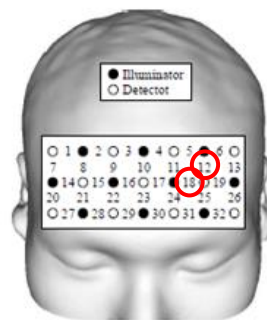


Fig5. Positions of optical fibers and channels

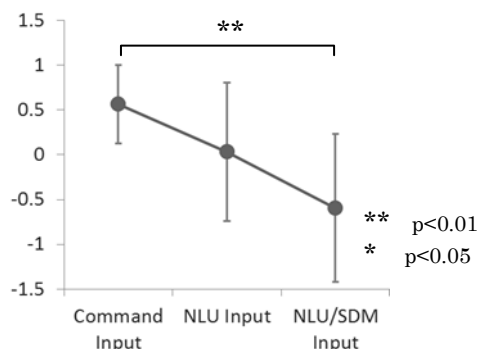


Fig6. Z-value of Oxy-Hb at left lateral portion (12ch)
(ANOVA, $p=0.0038$, $F=6.89$)

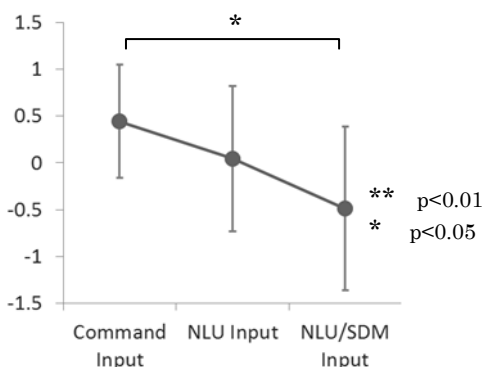


Fig7. Z-value of Oxy-Hb at left lateral portion (18ch)
(ANOVA, $p=0.0357$, $F=3.78$)

脳内 Hb は神経活動との相関が高い酸素化ヘモグロビン(Oxy-Hb)の変化に着目する[Hoshi 1998, 2001]. 呼吸や血圧の変動を除外するため 0.2Hz のローパスフィルタを掛けた後、発話中の Oxy-Hb 変化量から平均値を求め、標準得点化を行った。すべてのチャンネルごとに一要因分散分析を行ったところ、危険率 5% 以下で有意差が認められたのは 12ch と 18ch の Oxy-Hb 量(図 5 に示す丸印位置)の変化であり、共に前頭前野左外側の活動であった。これら仕様間の比較を図 6 と 7 に示す。多重比較による検定では、Command Input に比べ NLU/SDM Input で有意な減少が見られた。これは、音声入力による脳活動の低下を反映していると考えられ、特に左脳活動は言語機能に関わる部位とされることから、Command Input では、音声 IF へ受諾させるため、語彙や語順などを生成しようと脳が積極的に働いていることが考えられる。逆に NLU Input や

NLU/SDM Input では、言語生成の作業が減りスムーズな音声入力が行えた結果と推察できる。

5. 考察

音声 IF 使用時の一文あたりのコマンド入力数を確認してみると、Fig8 に示すように、文章に含まれるコマンド数の割合は、Command Input に比べ NLU Input で減少を示し、さらに NLU/SDM Input では、有意にコマンド数が少ない。これは Command Input では音声入力の際に、自分の意図する操作内容をシステムが受諾できるコマンドに変換して発話する必要があるため、コマンドに関する知識不足に起因する変換作業の誤りにより、コマンドの繰り返し、言い直しが発生し、コマンド数が増加したと考えられる。一方、NLU Input や NLU/SDM Input では、自分の意図する操作内容をそのまま発話しても、多くの場合、システムが受諾できるためコマンド数は少なく済む。

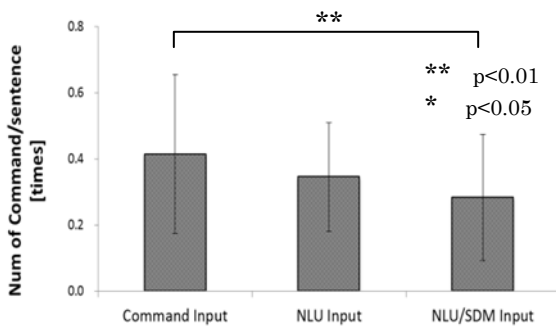


Fig8. Mean of command rate in sentence (ANOVA, $p=0.0021$, $F=6.37$)

以上のことを踏まえ脳科学の知見と生理指標の結果を比較する。発話に関係する脳機能が左脳ブローカー野にあり[辰巳 2004]、高次な情報処理は前頭葉で行われる[Carlson 2010]ことから、意図をコマンドに変換する作業が前頭左外側部で行われると仮定してみると、Command Input では 12ch, 18ch の Oxy-Hb 量を亢進させ、NLU/SDM Input では Oxy-Hb 量が抑制させた結果と対応付けができる。すなわち今回の検討で、自然言語理解や対話管理機能を搭載した音声 IF は、コマンドの変換作業に関わる高次な脳内作業の負担を低減させる一定の効果があり、Oxy-Hb 量の変化がこれを裏付けていると言える。

6. まとめ

車載音声 IF の対話デザインの知見を得るため、ドライバ負担の観点から、ドライバの生理変化の評価を行った。運転中の音声 IF 仕様の違いにおいて、自律神経系の指標である心電図と皮膚電位活動では差異は見られなかったが、中枢系の指標である脳血流に有意な変化があることが分かった。自然言語理解や対話管理などの機能を使うと、ドライバの前頭左外側部の脳血流量を下げ、言語情報処理への負担を緩和させていることが示唆される。これはドライバの音声 IF 入力の負担を低減させる効果を定量的に示した結果と考えられる。

今回検討したドライバ負担の観点は、車載音声 IF 知能化デザインの一助になると考える。車での音声 IF の使用は、従来の音声 IF が持つ機能要件の他、ドライバがいつ、どこで、使用するか、ドライバの負担になっていないかなどの要件が求められる。音声 IF の機能とドライバ負担の関係を明らかにすることで、負

担低減効果を持たせたデザイン設計や実用性について検討できる。

一方で、ドライバの負担感は、個人の経験や嗜好でも変わるため、ドライバの個人特性や状態も考慮に入れる必要がある。今後は、個人適応する音声 IF デザインについても検討していきたい。

参考文献

- [Hilton 2012] G. Hilton et al.: Deep Neural Networks for acoustic modeling in speech recognition, IEEE Signal Processing Mag, pp.82-97, 2012.
- [Sagae 2009] Kenji Sagae et, al.: Towards Natural Language Understanding of Partial Speech Recognition Results in Dialogue Systems, Proceedings of NAACL HLT 2009: Short Papers, pp.53-56, 2009.
- [神沼 2015 春] 神沼充伸, 松田佳子, 和氣早苗: 自然言語理解を用いた音声インタフェースの効果, 自動車技術会春季大会, 2015.
- [神沼 2015 秋] 神沼充伸, 清水俊行: 施設カテゴリ検索における車載型自然言語音声インタフェースの効果, 自動車技術会秋季大会, 2015.
- [Fraser 1991] N.M Fraser, G.N. Gilbert: Simulating speech systems. Computer Speech and Language, Vol.5, No.1, pp.81-99, 1991.
- [Bernsen 1994] N. Bernsen, et al.: Wizard of Oz Prototyping: How and when? CCI Working Papers in Cognitive Science and HCI, WPCS-94-1. Center for Cognitive Science, Roskilde University, 1994.
- [小西 2000] 小西郁夫, 竹内貞夫, 老川幸夫, 他: 近赤外光による無侵襲生体計測マルチチャンネル酸素モニタ OMM-2000 の開発; 島津評論, Vol.57, No.1・2, pp.141-151, 2000.
- [網田 2005] 網田孝司, 常石召一, 河野理, 他: 近赤外光イメージング装置 NIRStation の開発; 島津評論, Vol.61, No.3・4, 2005.
- [Hoshi 1998] Y. Hoshi, S. Kosaka, Y. Xie, et al.: Relationship between fluctuations in the cerebral hemoglobin oxygenation state and neuronal activity under resting conditions in man, Neuroscience Letters, 245, pp.147-150 1998.
- [Hoshi 2001] Y. Hoshi, N. Kobayashi, M. Tamura: Interpretation of nearinfrared spectroscopy signals: A study with a newly developed perfused rat brain model: Journal of applied physiology, Vol.90, No.5, pp.1657-1662, 2001.
- [大須賀 1993] 大須賀美恵子: 自律系生理指標を用いたメンタルワークロードの評価に向けて: 計測自動制御学会論文集, Vol.29, No.8, pp.979-986, 1993.
- [西村 1984] 西村千秋・小坂明生・常光和子・吉沢修治・南雲仁一: 自動車運転時における覚醒水準のフィードバック制御. バイオフィードバック研究, 11, pp.28-33, 1984.
- [辰巳 2004] 辰巳格: 音声知覚・理解・意味処理・音声生成にかかわる脳のネットワーク, 音声研究, 日本音声学会, 8(3), pp.15-27, 2004.
- [Carlson 2010] Neil R. Carlson: 神経科学テキスト 脳と行動, 第 3 版, 丸善株式会社, 2010.