

Assessing error recognition in automated driving

Spiessl, W., & Hussmann, H.

(2011). IET Intelligent Transport Systems, Vol. 5, Iss. 2, pp.103-111.

1 Introduction and Background

- 自動化システムは、人間が行う活動を遂行するテクノロジー
(Parasuraman & Riley, 1997)
- 特に transportation における、オペレータの作業は著しく軽減
 - 船, 飛行機の自動運転は既に存在 (Lee & Sanquist, 1996; Wiener & Curry, 1980)
 - 自動車の自動運転も可能になるだろう
- 自動車の自動運転によって、人間の運動や認知に関する資源は運転から解放される
 - 人間は運転とは関係のないタスク(電話, メールチェック等)を行うだろう
(Flemisch et al, 2008)
- しかし、自動化システムは完璧ではない
 - 自動化システムがエラーを起こした場合、人間が自動化システムの代わりを行う
 - つまり、人間は監視制御の役割を担う (Sheridan, 2002)
 - ◇ 監視制御では、自動化システムへの信用が重要な鍵となる
 - 自動化システムへの低い信用は、過度の未使用(Disuse)を招く
 - 自動化システムへの高い信用は、過度の使用(Misuse)を招く
(Parasuraman & Riley, 1997)
- 疑問点
 - 自動運転におけるエラー認識はどのように行われるか
 - 運転には関連のないサブタスクがどのようにエラー認識に影響するか

2 Methodology

2.1 Background

- 路線変更テスト(ALCT: Autonomous lane change test)を実験に使用

2.2 Functional principles and interaction

- ALCT の仮想運転環境
 - 自動運転で走行(毎時 60 キロ)
 - 路線変更を指示する標識が出現し、標識に合わせて自動的に路線変更が行われる
- 3つの自動運転エラー (Fig.1)

- error 1 : 直進標識が出現し, 路線を変更する
- error 2 : 路線変更標識が出現し, 直進する
- error 3 : 路線変更標識が出現し, 誤った路線に侵入する
- 自動運転のエラーが生じた場合, 手動でハンドルを操作して走行を修正

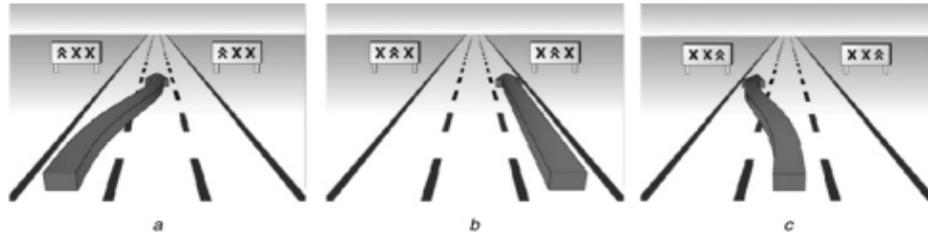


Fig. 1 Error types

- a Lane change when not required (error 1)
- b No lane change when required (error 2)
- c Change to the wrong lane (error 3)

2.3 Metrics

- 従属変数
 - 反応時間 : 自動運転エラーが生じてから, 手動でハンドルを切るまでの時間
 - missed error : 自動運転エラーを修正できなかった回数
 - false reaction : 正常に自動運転が行われているときに, ハンドルを切った回数

2.4 Equipment

- ALCT では, driving simulator を使用 (Fig. 2)
 - 前方に 2 つの走行画面, ハンドル横に 1 つのサブタスク画面



Fig. 2 Set-up of the ALCT with a car mock-up and plasma screens in a driving simulation environment

Picture displays error type 3

3 Formal categorization of secondary task

- 同一の資源によって 2 つのタスクが処理される場合, 処理の干渉が生じる

(Wickens & Liu, 1988)

➤ 自動運転の監視には視覚的資源が必要，サブタスクでも視覚的資源が必要

● サブタスクの分類

➤ Modality : 視覚/聴覚 (visual/auditory) (Wickens, 1992)

➤ Information coding : 言語/空間 (verbal/spatial) (Wickens, 1992)

➤ Degree of interaction : active/passive

◇ active : 操作的 - 話す

◇ passive : 消費的 - 読む, 聞く

➤ Interruptibility : good/bad (Alliance of Automobile Manufactures: AAM, 2006)

◇ どの程度メインタスクが妨害されているかを主観評定

● Table 1 は分類を行った結果

➤ 自動運転の監視を干渉する処理 : *visual – active – bad interruptibility – spatial*

Table 1 Categorisation of secondary tasks using different dimensions interfering with the primary driving task, along with examples, if found

| Modality | Degree of interaction | Interruptibility | Information coding | In-car task examples |
|-----------------|-----------------------|------------------|--------------------|---|
| <i>visual</i> | <i>active</i> | <i>good</i> | <i>verbal</i> | navigation, music selection |
| <i>visual</i> | <i>active</i> | <i>good</i> | <i>spatial</i> | interactive map |
| <i>visual</i> | <i>active</i> | <i>bad</i> | <i>verbal</i> | - |
| <i>visual</i> | <i>active</i> | <i>bad</i> | <i>spatial</i> | - |
| <i>visual</i> | <i>passive</i> | <i>good</i> | <i>verbal</i> | news headlines and SMS |
| <i>visual</i> | <i>passive</i> | <i>good</i> | <i>spatial</i> | follow navigation hints |
| <i>visual</i> | <i>passive</i> | <i>bad</i> | <i>verbal</i> | news article text |
| <i>visual</i> | <i>passive</i> | <i>bad</i> | <i>spatial</i> | - |
| <i>auditory</i> | <i>active</i> | <i>good</i> | <i>verbal</i> | speech commands |
| <i>auditory</i> | <i>active</i> | <i>bad</i> | <i>verbal</i> | talking on the phone |
| <i>auditory</i> | <i>passive</i> | <i>good</i> | <i>verbal</i> | listening to music |
| <i>auditory</i> | <i>passive</i> | <i>bad</i> | <i>verbal</i> | news, traffic information, audio book and audio memos |

4 Experimental assessment

4.1 Interaction versions

● LO version

➤ 自動運転中にハンドルは静止

◇ 手動ハンドル操作はバイナリー入力

◇ 左右に 90 度以上ハンドルを切る

● Hi version

➤ 自動運転中にハンドルは運転と連動

◇ 手動では 0.5 秒間 3Nm 以上の力を加える

◇ ハンドルと連動して自動車が走行

● 手動で路線変更後，自動的に自動運転に切り替わる

- 仮説
 - ハンドルからの直接的なフィードバックによって、Hi version で、自動運転エラーへの反応時間は短く、エラー修正できない(missed error)回数は少ない

4.2 Study 1 – LO version

- LO version でのみ実験を実施，サブタスクの影響を検討

4.2.1 Experimental procedure

- 日常的に行われるであろう 6 つのサブタスクを選択
 - navigation : 提示される目的地をカーナビに入力
(*visual – active – good interruptibility – verbal*)
 - interactive map : 提示される目的地を地図で探索
(*visual – active – good interruptibility – spatial*)
 - phone call : 電話で会話
(*auditory – active – bad interruptibility – verbal*)
 - audio book : audio book を聞く
(*auditory – passive – bad interruptibility – verbal*)
 - long text : 長文リーディング
(*visual – passive – bad interruptibility – verbal*)
 - short text : 短文リーディング
(*visual – passive – good interruptibility – verbal*)
- サブタスク
 - navigation, interactive map
 - ◇ 解答入力直後に次の問題が提示される
 - phone call
 - ◇ 4 択の一般常識問題を口頭で出題，解答をしてその理由を述べる
 - audio book, long text, short text
 - ◇ 各試行後に理解度テストを実施
- 参加者ごとに 7 試行実施
 - 各試行で各サブタスクを使用，1 試行はサブタスクなし(baseline)で実施
- 自動運転エラー
 - 1 試行に 90 回標識が出現，10 回中 1 回で自動運転がエラー
 - ◇ 3 種類のエラーが各 3 回(合計 9 回)ランダムな順序に生じる

- ヘッドマウントカメラを使用して視線を測定
- 各試行後に作業負荷の主観評定(DALI: Driver activity load index)を実施

4.2.2 Sample

- 参加者：28名（男性19名，女性9名，平均年齢29.4歳）

4.2.3 Results

反応時間 (Fig. 3)

- サブタスクの一要因分散分析
 - サブタスクの主効果がみられた
($F(3.644, 98.389)=9.545, p<0.001$, Greenhouse-Geisser correction)
 - ◇ Benferoni法を使用して一対比較
 - baseline < audio book 以外のサブタスク ($p<0.05$)
 - audio book < navigation ($p=0.023$), interactive map ($p=0.018$)
 - short text < navigation ($p=0.002$), interactive map ($p=0.010$)

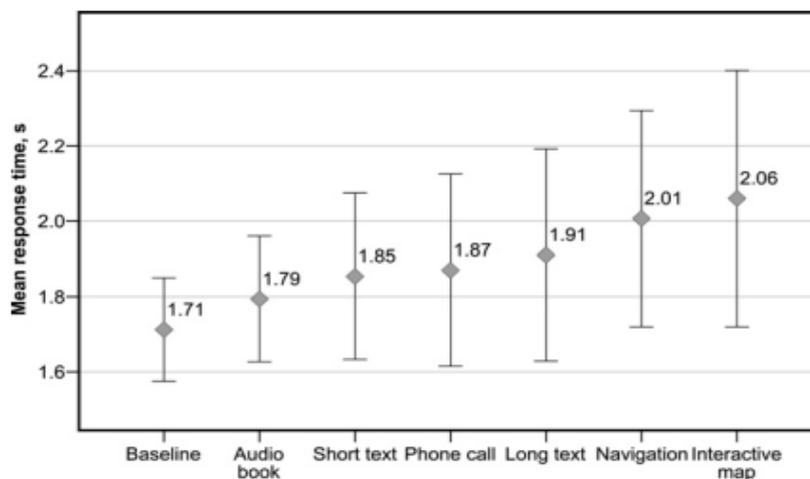


Fig. 3 Mean response time with standard deviation for the baseline and all secondary tasks (sorted in ascending order)

- サブタスクを degree of interaction に基づいて 3 つに分割
 - active(navigation, interactive map, phone call), passive(audio book, long text, short text), baseline で比較
 - degree of interaction の主効果がみられた
($F(1.645, 44.417)=26.220, p<0.001$, Greenhouse-Geisser correction)
 - ◇ Benferoni法を使用して一対比較
 - 全ての組み合わせで有意差 ($p<0.001$)

- baseline が最短, active のサブタスクが最長
- サブタスクを modality に基づいて 3 つに分割
 - auditory(audio book, phone call), visual(long text, short text, navigation, interactive map), baseline で比較
 - modality の主効果がみられた($F(2, 54)=19.361, p<0.001$)
 - ◇ Benferoni 法を使用して一対比較
 - auditory < visual ($p<0.013$)
- error type の一要因分散分析
 - error type の主効果がみられた($F(2, 54)=41.263, p<0.001$)
 - ◇ Benferoni 法を使用して一対比較
 - 全ての組み合わせで有意差
 - error 3($M=1.73$)が最短, error 2($M=2.09$)が最長

Missed errors (Fig. 4a)

- 全参加者の全試行で 64 回 missed error が生じた
 - audio book で最小, interactive map で最大
- missed error におけるサブタスクの影響がみられた($\chi^2(6)=73.414, p<0.001$)
- サブタスクを degree of interaction に基づいて分割
 - passive (4 回) < active (59 回) ($\chi^2(1)=50.104, p<0.001$)

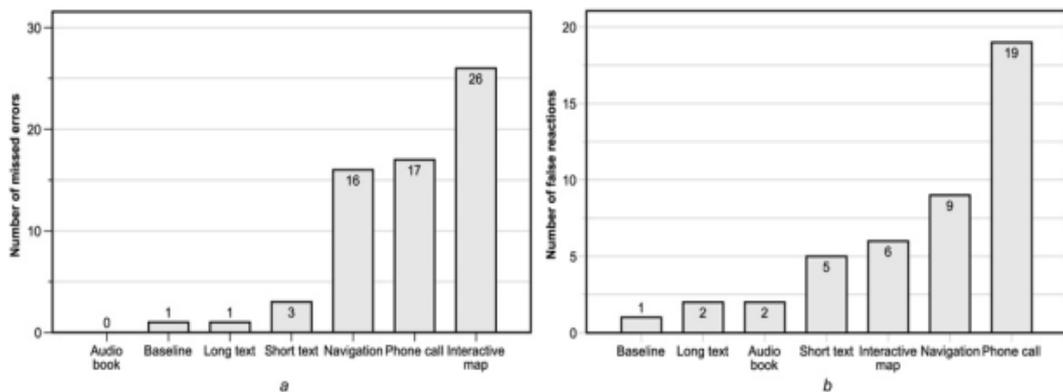


Fig. 4 In each condition summed over all participants (each sorted in ascending order)

a Number of missed errors

b False reactions

False reactions (Fig. 4b)

- 全参加者の全試行で 44 回 false reaction が生じた
 - baseline で最小, phone call で最大
- false reaction におけるサブタスクの影響がみられた($\chi^2(6)=38.413, p<0.001$)

Gaze behavior (Fig. 5)

- 1 回の注視時間を測定(200ms 以上視線を止めた場合のみ)
- 2 箇所画面に分けて分析
 - CID(central information display) : サブタスク画面
 - FRONT(driving scene) : 走行画面
- ☆ それぞれでサブタスク(visual)の一要因分散分析
 - CID でサブタスクの主効果がみられた
($F(3, 81)=44.210, p<0.001$; Green-Geisser correction)
 - interactive map で最短, long text で最長
 - FRONT でサブタスクの主効果がみられた
($F(2.283, 61.650)=53.455, p<0.001$; Green-Geisser correction)
 - long text で最短, short text で最長

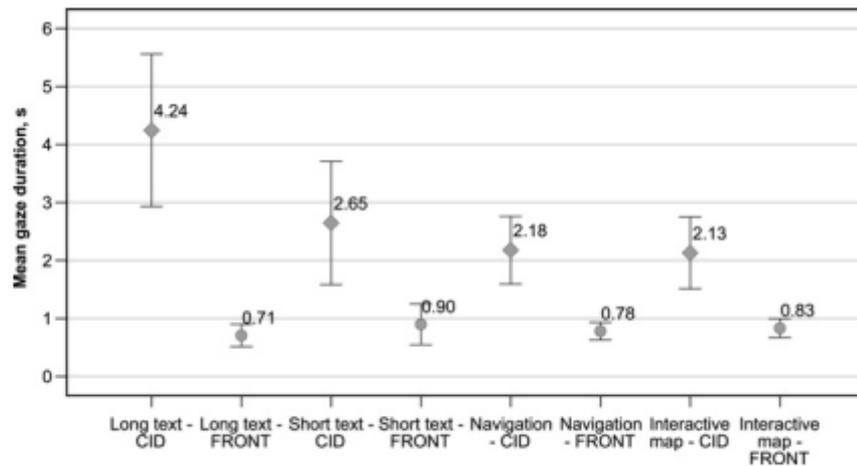


Fig. 5 Mean gaze duration on AOIs and standard deviation (in s)

Subjective rating

- 作業負荷の 6 段階評定(0=low~5=high)
 - audio book で最小($M=1.71$), interactive map で最大($M=3.18$)

4.2.4 Summary and discussion of result in study 1

- サブタスクは, 自動運転の監視パフォーマンスに影響

- active なサブタスク, visual のサブタスクは, 自動運転の監視を妨げる
- error type は反応時間に影響
 - error 2 で反応時間は最長
 - ◇ 動きを伴わないエラー(omission error)が原因だろう
- サブタスク画面(CID)を長く見ているほど, missed error の回数が少ない
 - interactive map では, 合計 26 回の missed error が生じた
 - その内 18 回で, 参加者は走行画面(FRONT)を見ていたにも関わらずエラー修正で
きなかった
 - ◇ 作業負荷が高い場合, 走行画面を見ていても, 自動運転エラーを適切に修正
できない

4.3 Study 2 – LO version against HI version

4.3.1 Experimental procedure

- 実験 1 から 3 つのサブタスクを使用
 - interactive map, short text, phone call
- 基本的な手続きは実験 1 と同様
 - LO version と HI version で各サブタスクを実施
 - baseline は HI version で実施
 - 参加者ごとに 7 試行(2(version)×3(サブタスク)+1(baseline))実施

4.3.2 Sample

- 参加者 : 29 名 (男性 20 名, 女性 9 名, 平均年齢 34.2 歳)

4.3.3 Results

反応時間 (Fig. 6)

- 2(version)×3(サブタスク)×3(error type)の分散分析
 - version の主効果がみられた($F(1, 28)=133.64, p<0.001$)
 - ◇ Hi version < LO version
 - サブタスクの主効果がみられた($F(2, 56)=15.14, p<0.001$)
 - ◇ サブタスクについて Benferoni 法を使用して一対比較
 - interactive map > phone call($p=0.001$), short text($p=0.001$)
 - error type の主効果がみられた
($F(1.32, 36.88)=38.42, p<0.001$, Greenhouse-Greisser correction)

- ◇ error type について Benferoni 法を使用して一対比較
 - 全ての組み合わせで有意差 ($p < 0.001$)
 - error 3 が最短, error 2 が最長
- version と error type の交互作用
 - ◇ LO version の時に, error1, 2 と error 3 との差が顕著

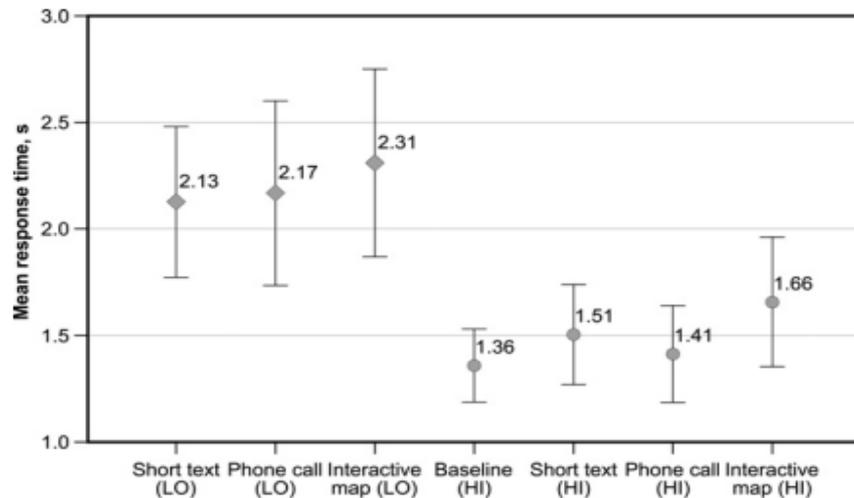


Fig. 6 Mean response time with standard deviation for all tasks performed using the LO version and the HI version

- 年齢と反応時間との間に相関関係が認められた (Table 2)
 - 年齢が高いほど反応時間が長い

Table 2 Correlations response times to age (Pearson-r)

| | Short text (LO) | Phone call (LO) | Interactive map (LO) | Short text (HI) | Phone call (HI) | Interactive map (HI) |
|--------------------|----------------------|----------------------|----------------------|---------------------|----------------------|----------------------|
| correlation to age | 0.57 ($p = 0.001$) | 0.51 ($p = 0.005$) | 0.4 ($p = 0.03$) | 0.26 ($p = 0.17$) | 0.42 ($p = 0.023$) | 0.34 ($p = 0.076$) |

Missed errors (Table 3)

- LO version の全試行で 55 回, HI version の全試行で 33 回 missed error が生じた
 - version 間, 各 version 内で有意差はみられなかった

False reactions

- HI version で false reaction の数が多い(190 回)

Table 3 Number of missed errors and false reactions in each condition summed over all participants in study 2

| | Short text (LO) | Phone call (LO) | Interactive map (LO) | Short text (HI) | Phone call (HI) | Interactive map (HI) | Baseline (HI) |
|-----------------|-----------------|-----------------|----------------------|-----------------|-----------------|----------------------|---------------|
| missed errors | 10 | 24 | 16 | 7 | 11 | 13 | 2 |
| false reactions | 1 | 11 | 7 | 24 | 74 | 34 | 58 |

Gaze behavior

- CID(central information display) : サブタスク画面
- 1回の注視時間を測定(200ms以上視線を止めた場合のみ)
 - 2(version:LO/HI)×2(サブタスク:interactive map/short text)の分散分析
 - ◇ versionの主効果($F(1, 25)=12.09, p<0.01$)
 - HI version < LO version
 - interactive map-LO: $M=2.22s$, HI: $M=2.00s$
 - short text-LO: $M=2.52s$, HI: $M=2.29s$

4.3.4 Summary and discussion of result in study 2

- HI versionで反応時間は短く, missed errorは少ない
 - ハンドルからのフィードバックの影響
 - LO versionでは, ハンドルからのフィードバックはなく, 警戒を保つ必要がある
 - ◇ 作業負荷が大きくなり, ドライバーにネガティブな影響を与える
- HI versionでサブタスク画面を見る時間が短い
 - HI versionの方が走行画面に注意を向けている
- 我々は, HI versionを勧めるが, HI versionでは false reactionが生じる危険がある

5 Conclusion

- 新たな実験手法を提示
- サブタスクが監視制御に強い影響を及ぼす
 - degree of interaction(active/passive), modality(auditory/visual)の違いが顕著