音声言語情報処理 32−17
（2000．7．15）

運転中における音声対話システムの評価

清水 司　小島 真一　脇田 敏裕　本郷 武朗
（株）豊田中央研究所
shimizu@iclab.tytlabs.co.jp, skojima@mosk.tytlabs.co.jp,
toshi@dii.tytlabs.co.jp, hongo@dii.tytlabs.co.jp

あらまし
車載用音声対話システムの評価について述べる。運転中に音声対話システムを用いて交通情報検索タスクとLED刺激に対する反応タスクを同時に行い、タスク達成時間と反応時間を測定した。タスク達成時間からシステムの利便性の評価を行い、LED刺激に対する反応遅れの割合からシステムの安全性の評価を行った。また、異なる対話方式に対してタスク達成時間のモデル化を行い、その有効性についても論じる。
キーワード
音声対話システム、タスク達成時間、反応遅れ、安全性、利便性

Evaluation of Spoken Dialog Systems for a Vehicle

Tsukasa Shimizu　Shin'ich Kojima　Toshihiro Wakita　Takero Hongo
Toyota Central R&D Labs., Inc.
shimizu@iclab.tytlabs.co.jp, skojima@mosk.tytlabs.co.jp,
toshi@dii.tytlabs.co.jp, hongo@dii.tytlabs.co.jp

Abstract
We describe the evaluation of spoken dialog systems for a vehicle. While driving, subjects performed the dual task which consisted of the traffic information retrieval by using spoken dialog systems and the detection for a LED stimulus. Their total task time for the traffic information retrieval and their detection time were measured. Usability and safety of the systems were evaluated by the total task time and by the rate of the delayed detection time, respectively. We also propose the models of the total task time for the different dialog control strategies and discuss their effectiveness.
key words
spoken dialog, total task time, delayed reaction time, safety, usability

1 はじめに

近年、カーナビなどの車載情報機器の操作では、音声対話入力への関心が高まっており、カーナビを対象とした対話システムの研究も行われている[1]。音声対話入力は、画面やボタンへの注視やハンドルから手を離す必要がないなどの点で、手操作よりも安全であると期待されている。しかしながら、現在、運転中に適した音声対話方式は明確ではなく、さらに、その評価法も確立されていない。

我々は、車載用音声対話システムでは、主に利便性と安全性の2つの観点から評価することが重要であると考えている。また、実際に対話システムを開発する際には、試作・実車実験を行い、評価を行っているのが現状であり、設計段階でシステムの評価を行えることが望ましい。

我々は、手操作のインタフェース評価に用いられている構成的手法[2]を参考に、利便性と安全性の観点から、音声対話システムを設計段階で評価する手法の開発を進めている。構成的手法とは、インタフェースの構成部品毎に視認、判断、操作などにかかる時間や回数などの評価尺度をモデル化し、それらの組合せによって、インタフェース全体を評価する手法である。

今回、利便性の尺度としてタスク達成時間、安全性の尺度としてLED刺激に対する反応遅れの割合を用いて対話システムの評価手法の検討を行った。3つの対話方式に対して、運転中に交通情報検索タスクと、LED刺激に対する反応タスクを同時に行い、それぞれの対話方式でのタスク達成時間とLED刺激への反応遅れ割合を比較した。また、タスク達成時間に関しては、各対話方式での構成部品（条件入力や確認応答）の時間からタスク達成時間を推定するモデルを作成し、その検証を行った。

2　利便性と安全性

2.1　利便性とタスク達成時間

車の中で音声対話システムを用いた交通情報検索を想定した場合、得られる情報はタイムリーである必要がある。例えば、進路上の交通情報を検索するとしよう。対話に時間がかかり、渋滞を避けるために曲がるべき交差点を過ぎてから、「この先渋滞しています」との情報を得たとしても有用ではない。このような場合、多少、発話の自由度や自然性が損なわれたとしても、できるだけ早くタスクを達成できることが重要である。そこで今回、利便性の尺度としてタスク達成時間を用いた。

2.2　安全性とLDE刺激への反応遅れ割合

音声対話の安全性は、音声対話が運転に与える影響として捉えることができる。文献[3]では、ドライバに対して記憶課題を課した場合には、課していない場合に比べて、運転席前方に設置されたLED刺激に対して反応遅れの生じる頻度が増加すると報告している。そこで今回、この手法を用いて安全性の尺度として、音声対話中のLED刺激に対する反応遅れの割合を用いた。

3　実験

3.1　実験装置

実験は，音声対話システムとLED刺激提示・反応収集システムを搭載した実験車で行った。実験中の様子を記録するために、実験車の前方風景と被験者の視線をVTRに収録した。

対話システムはLANで接続された2台のPCで構成されており、1台は音声認識、対話制御および交通情報検索を行い、もう1台は対話応答文および検索結果を音声合成で読み上げる。音声認識エンジンには（株）国際電気通信基礎技術研究所のATRSPRECを用い、音声合成エンジンにはの沖電気工業（株）のSmart Talkを用いた。また対話システムは、ドライバ発話の認識結果とシステムの応答文のログを時間情報とともに記録する。

図1にLED刺激提示・反応収集システムの構成を示す。上下に接近したLED対をドライバの周辺視に相当する位置に左右一組ずつ設置し、それらの点灯をPCで制御する。また、ドライバの反応は、シフトレバーに取り付けたスイッチによりPCで記録する。

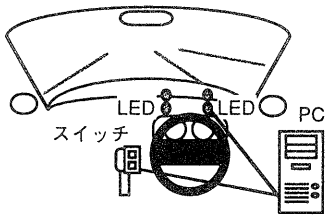

図1　LED刺激提示・反応収集システム構成

3.2　タスク

被験者は、1周約8分の社内の巡回コース（右左折12回と緩いカーブが4回含まれる）を1周する間に、LED刺激反応タスクと交通情報検索タスクを同時に繰り返し行う。

交通情報検索タスクでは、被験者は後述する対話方式に従って、予め決定した時刻と経路（始点と終点地名で指定）の交通情報検索を行う。1回の検索が終了すると3秒後に自動的に次の検索が始まるようにした。

LED刺激反応タスクでは、図1に示した4つLEDの内のいずれか1つが点灯し、被験者は点灯したLEDの上下の判定をスイッチによって答える。LED対を2組用意したのは、被験者の注意を特定のLED対に集中させないためである。LEDの点灯間隔は、平均が6秒となるように5秒、6秒、7秒でランダムとした。LEDの点灯後2秒経っても被験者の反応がない場合は見落としとみなした。

3.3　音声対話方式

今回設定した交通情報検索は、時刻と経路の始点地名、終点地名の3つの検索条件の入力を必要とするものである。これらの検索条件の対話方式として、Slot Filling方式、定型文方式および単語列挙方式の3つの比較を行った。それぞれの方式の対話例を図2に示す。Slot Filling方式（図2-A）は、システムの問いかけに従って、3つの検索条件を1つずつ入力する最もオーソドックスな対話方式である。定型文方式（図2-B）は、定型の文型であるが、1度に3つの検索条件を文の形で入力する対話方式である。単語列挙方式は（図2-C）は、対話制御方式が定型文方式と同じであるが、検索条件を文の形ではなく、単語の羅列で入力する方式である。これらの3つの対話

A　Slot Filling方式

sys：渋滞情報を検索します。
sys：時刻を言って下さい。
drv：９時
sys：９時ですね？
drv：いいえ
sys：時刻を言って下さい。
drv：３時
sys：３時ですね？
drv：はい
sys：起点を言って下さい。
drv：中研
sys：中研ですね？
drv：はい
sys：終点を言って下さい。
drv：藤ヶ丘
sys：藤ヶ丘ですね？
drv：はい
sys：３時ごろの中研から藤ヶ丘までの渋滞情報を検索します。
sys：渋滞しています。

B　定型文方式

sys：渋滞情報を検索します。
sys：検索条件を言って下さい。
drv：１０時ごろの本郷から大須まで
sys：１０時ごろの本郷から大須までですね？
drv：はい
sys：１０時ごろの本郷から大須までの渋滞情報を検索します。
sys：渋滞しています。

C　単語列挙方式

sys：渋滞情報を検索します。
sys：検索条件を言って下さい。
drv：３時 長久手 豊田
sys：３時ごろの長久手から豊田までですね？
drv：はい
sys：３時ごろの長久手から豊田までの渋滞情報を検索します。
sys：渋滞しています。

図2　各音声対話方式での対話例
sysはシステム発話、drvはドライバ発話である。

方式に共通して検索条件の入力後にシステムから確認応答がある。それに対してドライバ（被験者）は、システムが正しく検索条件を認識している場合に「はい」、誤認識している場合に「いいえ」で答える。「いいえ」と答えた場合には、システムは再度、検索条件の入力を促す。今回用いた認識エンジンは何も認識しない（不認識）場合があり、その際にはシステムは「もう一度言って下さい」と入力を促す。

これらの3つの対話方式のうち、Slot Filling方式はタスク達成時間が長いが、システムとの対話のやり取りが単純なため、運転に対しては低負荷（影響が少ない）と予想される。一方、定型文方式はタスク達成時間は短いが、それぞれの検索条件から文を構成するという思考過程があるため、運転に対して何らかの影響がでる可能性がある。また、単語列挙方式は、定型文方式に比べて、タスク達成時間の点では同等であるが、文構成という思考過程がないため、運転に対する影響が異なる可能性がある。

3.4　被験者と実験条件

被験者は社内の30代の男性3人（S1,S2,S3）とした。1回の実験では、(1)LED刺激反応タスクのみ（音声対話なし条件）、(2)Slot Filling方式による交通情報検索タスク＋LED刺激反応タスク（Slot Filling条件）、(3)定型文方式による交通情報検索タスク＋LED刺激反応タスク（定型文方式条件）、(4)単語列挙方式による交通情報検索タスク＋LED刺激反応タスク（単語列挙方式条件）の4つの条件を連続的に行った。また、1回の実験の中では、3つの対話方式ともに交通情報検索タスクの条件は同じにした。実験はそれぞれの被験者に対して数日間かけて、条件の順番を変えながら10回行った。

4　実験結果

4.1　タスク達成時間

図3に、それぞれの対話方式におけるタスク達成時間の平均値と標準偏差を示す。全ての被験者で、タスク達成時間が最も長いのがSlot Filling方式であり、定型文方式と単語列挙方式では、同程度であった。同じ対話方式でも被験者間でタスク達成時間が異なるのは、主に認識率（検索条件入力における発声単位）の違いにより、検索条件入力と確認応答の回数が異なるからである。

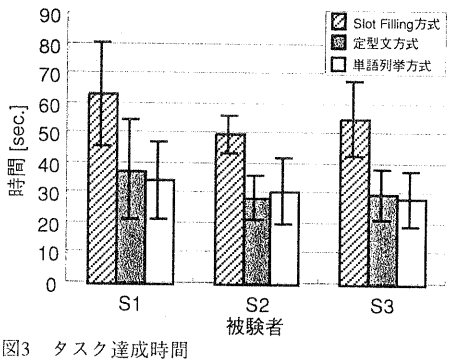

図3　タスク達成時間

表1に、検索条件入力における発声単位での認識率と誤認識率を示す。ここで、発声単位とは、Slot Filling方式では1つの検索条件単語であり、定型文方式では検索入力文の全体あり、単語列挙方式では3つの検索条件単

語の連なりである。表1中の認識率と誤認識率を足しても1.0にならないのは、不認識の場合があるからである。

表1　条件入力の認識率

被験者	Slot Filling方式		定型文方式		単語列挙方式	
	認識率	誤認識率	認識率	誤認識率	認識率	誤認識率
S1	82.8	12.1	63.5	12.2	77.0	13.5
S2	94.3	5.0	93.7	3.2	81.0	14.3
S3	94.3	5.7	84.7	3.5	92.1	3.4

表中の数字の単位は [%]

4.2　LED刺激への反応遅れ割合

反応遅れ割合は、実験の様子を収録したVTRの解析により、有効な反応時間のデータのみを用いて算出した。ここで、反応遅れは、各条件での”平均反応時間＋3σ”以上の反応時間と定義した。σは、被験者毎に別途収集した停止時の音声対話なし条件の反応時間の分布の標準偏差である。図4に反応遅れの割合を示す。全被験者で音声対話なしの場合に最も反応遅れ割合が小さく、音声対話が運転に何らか影響を与えている可能性がある。

一方、対話方式間で比べると、予想とは異なり、Slot Fillig方式が他の方式に比べて必ずしも運転への負荷が低いものではなかった。むしろ被験者S3では、他の方式よりも負荷が高かった。また、定型文方式と単語列挙方式には明確な違いは見られず、検索条件の言い方の違いによる運転への影響は少ないと考えられる。

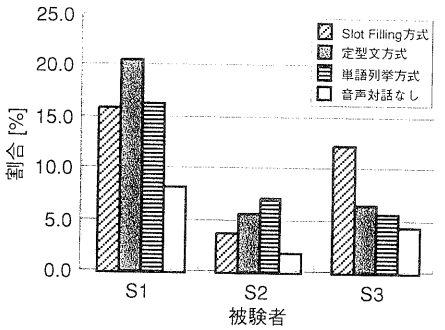

図4　LED刺激に対する反応遅れの割合

4.3　音声対話方式の評価

タスク達成時間と反応遅れ割合の2つの観点から対話方式間の評価を行った。横軸に反応遅れの割合、縦軸にタスク達成時間をとり、プロットしたものを図5に示す。図の中で原点に近いほど、タスク達成時間が短く、反応遅れの割合が小さい、すなわち、利便性と安全性が高いと見なせる。今回の実験からは、少なくともSlot Filling方式が運転中に適した音声対話方式であるとは言えない。また、定型文方式と単語列挙方式では顕著な違いは見られず、どちらが運転により適しているかは判断できない。

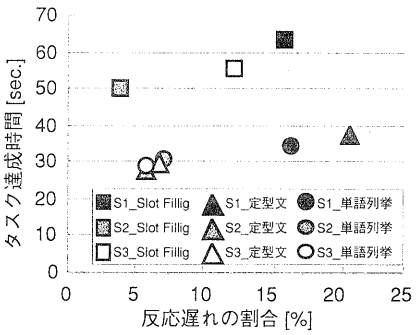

図5　音声対話方式の評価

5　構成的手法の検討

5.1　タスク達成時間の推定

実験に用いた各対話方式では、条件入力に対して必ず確認を行っており、対話の状態はシステムの音声認識率に従って遷移することになる。このような場合、対話に必要な時間（タスク達成時間）は、対話の各状態でのシステム発話の時間、ドライバの発話時間およびシステムの音声認識率から推定することができる。文献[4]では、対話の状態間の遷移確率を音声認識率で定め、対話の効率（発話交換数）について定量的に考察を行っている。

ここでは、文献[4]の手法をタスク達成時間に適用し、実験で得られた結果との比較を行う。文献[4]では、確認に対する応答（「はい（いいえ）」）は必ず認識されると仮定しているが、確認応答に対しても誤認識や不認識を起こすことを考慮して、推定式の導出を行った。

・タスク達成時間の推定式の導出

対話の状態を $S(i,j)$ で表し、1回の検索条件入力における対話の状態遷移を図6に示す。i はシステムが既知である（認識している）が、確認をしていない検索条件入力の数である。j はシステムが確認した検索条件入力の数である。1回の条件入力に対して考えているので、i および j は0または1であるが、今回のシステムでは不認識の場合もあるので、この場合を $i=N$ とする。

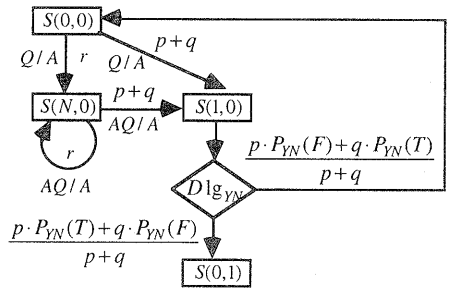

図6　条件入力対話における状態遷移

また、図6中のDlg_{YN}は後述する入力に対する確認対話であり、「はい」と認識した場合には$S(0,1)$に遷移し、「いいえ」と認識した場合には$S(0,0)$に遷移する。1回の条件入力おける対話の目標は、状態$S(0,0)$から状態$S(0,1)$に至ることである。

ここで、状態$S(i,j)$から状態$S(0,1)$に至るまでの時間の期待値を$T(i,j)$とすると、

$$T(0,0)=(p+q)\cdot\{T_{Q/A}+T(1,0)\}+r\cdot\{T_{Q/A}+T(N,0)\} \quad (1)$$

$$T(1,0)=\frac{p\cdot P_{YN}(T)+q\cdot P_{YN}(F)}{p+q}\cdot T_{YN} + \frac{p\cdot P_{YN}(F)+q\cdot P_{YN}(T)}{p+q}\cdot\{T_{YN}+T(0,0)\} \quad (2)$$

$$T(N,0)=(p+q)\cdot\{T_{AQ/A}+T(1,0)\}+r\cdot\{T_{AQ/A}+T(N,0)\} \quad (3)$$

となる。ただし、p、q、rはそれぞれ条件入力発話に対する認識率、誤認識率、不認識率である。$T_{Q/A}$は条件入力の際のシステムとドライバの発話交換（sys:「＊＊＊を言って下さい」　drv:「＊＊＊」）に要する時間であり、$T_{AQ/A}$は不認識後のシステムとドライバの発話交換（sys:「もう一度言って下さい」　drv:「＊＊＊」）に要する時間である。また、T_{YN}は確認対話Dlg_{YN}に要する時間である。$P_{YN}(T)$、$P_{YN}(F)$はそれぞれシステムが確認応答を正しく認識する確率と誤って認識する確率である。式(1)～式(3)より、1つの条件入力対話にかかる時間は次のように求まる。

$$T(0,0)=\frac{(p+q)\cdot(T_{Q/A}+T_{YN})+r\cdot T_{AQ/A}}{p\cdot P_{YN}(T)+q\cdot P_{YN}(F)} \quad (4)$$

図7に確認対話Dlg_{YN}での状態遷移を示す。確認対話における状態を$S(i)$で表す。iは確認応答に対するシステムの認識状態であり、0は何も認識していない状態であり、1は誤認識を含めて確認応答を認識した状態（である。また、$i=N$は不認識した状態とする。

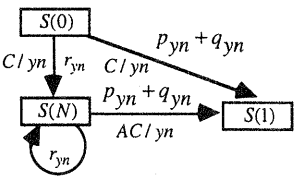

図7　確認対話における状態遷移

ここで、状態$S(i)$から最終状態$S(1)$に至るまでの時間の期待値を$T(i)$とすると、

$$T_{YN}=T(0)=(p_{yn}+q_{yn})\cdot T_{C/yn}+r_{yn}\cdot\{T_{C/yn}+T(N)\} \quad (5)$$

$$T(N)=(p_{yn}+q_{yn})\cdot T_{AC/yn}+r_{yn}\cdot\{T_{AC/yn}+T(N)\} \quad (6)$$

となる。ただし、p_{yn}、q_{yn}、r_{yn}はそれぞれ確認応答に対する認識率、誤認識率、不認識率である。$T_{C/yn}$は確認の際のシステムとドライバの発話交換（sys:「＊＊＊ですね？」　drv:「はい（「いいえ」）」）に要する時間であり、$T_{AC/yn}$は不認識後のシステムとドライバの発話交換（sys:「もう一度言って下さい」　drv:「はい」（「いいえ」））に要する時間である。式(5)、式(6)より、確認対話にかかる時間は次のように求まる。

$$T_{YN}=T_{C/yn}+\frac{r_{yn}}{p_{yn}+q_{yn}}\cdot T_{AC/yn} \quad (7)$$

また、確認対話で、最終的に確認発話（「はい（いいえ）」）を正しく認識する確率$P_{YN}(T)$、誤って認識する確率$P_{YN}(F)$は次のように近似できる。

$$P_{YN}(T)=\frac{p_{yn}}{p_{yn}+q_{yn}} \quad (8) \qquad P_{YN}(F)=\frac{q_{yn}}{p_{yn}+q_{yn}} \quad (9)$$

よって、式(7)～式(9)を用いて式(4)から、1つの条件入力対話における時間の推定式が求まり、各対話方式でのタスク達成時間は次のように導出される。

［Slot Fillig方式］

$$T=T_{Hello}+\frac{(p+q)\cdot(T_{Q/A}+T_{YN})+r\cdot T_{AQ/A}}{p\cdot P_{YN}(T)+q\cdot P_{YN}(F)}*3+T_{Srch}+T_{Ans}$$

［定型文方式、単語列挙方式］

$$T=T_{Hello}+\frac{(p+q)\cdot(T_{Q/A}+T_{YN})+r\cdot T_{AQ/A}}{p\cdot P_{YN}(T)+q\cdot P_{YN}(F)}+T_{Srch}+T_{Ans}$$

Slot Filling方式では、3回の検索条件入力があるので、式(4)の右辺を3倍する必要がある。また、上の式で、T_{Hello}、T_{Srch}、T_{Ans}はそれぞれ開始時のシステム発話、交通情報の検索、検索結果のシステム発話にかかる時間である。

・実測値との比較

ここでは、タスク達成時間の推定値と、実験から得られたタスク達成時間を比較し、推定式の検証を行う。

推定式中の時間パラメータ（$T_{Q/A}$，　$T_{AQ/A}$，　$T_{C/yn}$，$T_{AC/yn}$，T_{Hello}，T_{Srch}，T_{Ans}）は、実験で得られた3人の被験者の平均値を用いる。また、認識率パラメータ（p，q，r，p_{yn}，q_{yn}，r_{yn}）のうち、確認対話におけるものは、被験者間でほとんど違いがなかったので、全被験者内での認識率$\overline{p_{yn}}$、誤認識率$\overline{q_{yn}}$、不認識率$\overline{r_{yn}}$を用いる。さらに、各対話方式における検索条件入力の不認識率rは全被験者での不認識率$\bar{r}$を用い、誤認識率qは$1-p-\bar{r}$とする。図8に各対話方式の認識率とタスク達成時間の推定値との関係を示す。図8には、実験から得られたタスク達成時間も併せて示す。

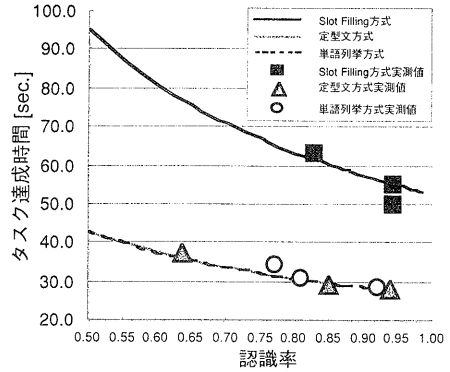

図8　タスク達成時間の推定

概ね±2秒以内の範囲でタスク達成時間を推定できている。また、全対話方式、全被験者から求めた、実測値と推定値の相関係数は、0.983である。したがって、音声対話システムの設計時に、おおよそのシステムおよびドライバの発話時間と、システムに用いる音声認識エンジンの認識率が分かれば、タスク達成時間の期待値を推定することが可能であり、対話方式間の比較ができる。

5.2　LED刺激への反応遅れ割合

以前、我々は、フリーリコールタスク（読み上げられた無関連な5つの単語を記憶し、それらを答えるタスク）中のLED刺激への反応遅れの分布を調べた。ある被験者の結果を図9に示す。図の横軸はタスクの経過時間であり、縦軸は反応遅れの頻度である。また、図中の縦線は、フリーリコールの記憶する区間と再生する区間の境界を示している。この被験者の場合、記憶区間よりも記憶を再生して答える区間で、反応遅れが多いことが分かる。特に、2つめの単語を答えるタイミングに反応遅れが多い。このような傾向は、他の被験者でも同様であった。

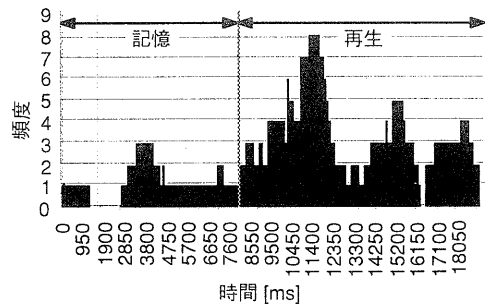

図9　フリーリコール中の反応遅れ分布

LED刺激への反応遅れ割合については、今回の実験結果からは対話方式間および被験者間に明確な傾向は見られなかった。しかし、フリーリコールで見られたように、対話中のドライバの各状態、例えば、システムからの質問を聞いている状態、検索条件を入力（発声）している状態、または確認応答の内容を吟味している状態などで、反応遅れがどのように分布しているかを解析することで、各対話方式について反応遅れ割合をモデル化できることが考えられる。今後、音声対話中の反応遅れの分布を詳細に分析する予定である。

6　おわりに

車載用の音声対話システムの対話方式について、利便性（タスク達成時間）と安全性（LED刺激への反応遅れ割合）の観点から評価を行った。今回用いた3つの対話方式（Slot Fillig方式、定型文方式、単語列挙方式）の中では、少なくとも、他の方式に比べてSlot　Filling方式が運転中に適した音声対話方式であるとは言えなかった。定型文方式と単語列挙方式では、明確な違いが見られなかった。

さらに、タスク達成時間をモデル化し、その有効性の検証を行った。作成したモデルは、概ね±2秒以内の範囲で実測値を推定することができ、その有効性を示すことができた。また、LED刺激への反応遅れ割合に関しては、今後データを詳細に分析することによってモデル化を図れる可能性を示唆した。

今回、LED刺激への反応遅れ割合のモデル化は行えなかったが、提案した評価手法によって、設計段階での音声対話システムの評価が期待できる。

参考文献

[1] 河野恭之, 屋野武秀, 笹島宗彦 : カーナビ音声対話システムMINOSの試作, 人工知能学会研究会資料, SIG-SLUD-9901-4, pp.21-26, (1999)

[2] T.Wakita and R.Terashima : Visual Behavior of navigation system operation while driving, Proceedings fo 6th World Congress on ITS, (1999)

[3] 小島真一他:音声対話の運転への影響評価法の開発, 自動車技術会学術講演会前刷集, N0. 91-99, pp.17-20, (1999)

[4] 新美康永, 西本卓也, 荒木雅弘 : 確認対話の制御方式の効率と音声認識システムの性能との関係, 情報処理学会研究報告, 99-SLP-27, pp.111-118, (1999)