

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 2 日現在

機関番号：15301

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26730092

研究課題名(和文)自動車運転状況の自動推定に基づく安全性に配慮した音声対話システムの研究

研究課題名(英文) Study on spoken dialogue system with safety consideration based on automatic estimation of driving situation

研究代表者

原直(Hara, Sunao)

岡山大学・自然科学研究科・助教

研究者番号：50402467

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：以下を実施した。

(1) 運転者の体や車両本体にとりつけたセンサから得られる生体信号や運転情報信号を用いて運転負荷の推定。スマートフォンで得られるセンサの利用を想定した運転負荷推定。(2) ユーザ負荷の観点で運転中の利用に適した対話戦略を評価。対話の主導権の方式と確認発話の方式を組み合わせた客観評価。(3) 対話システムとして評価するための主観評価。特に、運転を模擬した動作と音声対話システムの利用という二重タスク課題における主観評価。(4) ユーザ負荷の推定結果を組み込める柔軟な対話戦略の選択方式を実装。グラフ探索に基づく対話戦略を導入。客観評価実験と主観評価実験による性能評価。

研究成果の概要(英文)：We conducted below:

(1) We used bio signals and driving information signals that are measured by the driver's body and the vehicle's body for estimating the driving road from the signals. Furthermore, we estimated driving load assuming use of sensors obtained by smartphones. (2) We evaluated the spoken dialog strategy from the viewpoint of user's driving load. An objective evaluation by computer simulation was conducted by considering both dialog initiatives and the exitance of confirmation utterances. (3) We conducted a subjective evaluation to evaluate the performance of the proposed dialog strategy as a spoken dialog system. The subjects were asked to drive a simulator during talking with a spoken dialog system. (4) A dialog strategy based on graph search was introduced to realize a dialog strategy which considering the estimation result of the user's mental load. We evaluated the proposed system by objective evaluation and subjective evaluation.

研究分野：総合領域

キーワード：音声対話 情報検索 運転行動 生体信号 スマートフォン

## 1. 研究開始当初の背景

音声対話システムとは人とコンピュータが数回の音声のやり取りを行うことで様々な処理ができるシステムである。音声対話システムを構築する際には、「ドメイン」、「タスク」、「対話戦略」の3つを考える必要がある。ドメインとは知識源、タスクとはそのシステムで実行したい処理、対話戦略はタスク達成のための手順や方策である。例えば、バスの情報案内であれば、ドメイン=バスの運行情報、タスク=適切なバスを探す、であり、都市の観光案内であれば、ドメイン=都市の観光情報、タスク=観光地の情報を得る、となる。対話戦略はシステムにより様々な最適化が行われるが、一般には発話数を少なくすることで利用者の負担を軽減する手順を目指すことが多い。また、人との対話と同じような自然な対話を目指すこともある。

ユーザインタフェースの観点で言えば、タスクやドメインはバックエンドシステムの要素であり、対話戦略はフロントエンドシステムの要素である。従ってユーザのあらゆる行動が対話戦略への入力となり得る。マルチモーダルと呼ばれるシステムでは、ユーザからの能動的な入力を利用することが多く、ユーザの利用状況のような受動的な情報はあまり使われていない。そこで、本研究では、音声以外の様々なセンサ情報を活用して利用状況を推定する。そして、利用状況を考慮した音声対話システムの実現を目指す(図1)。

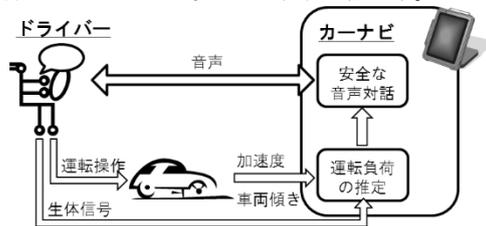


図1 研究の目的の概略図

## 2. 研究の目的

(1) スマートフォンのセンサ信号を利用して運転者の状況を推定する方式を検討する。利用者のシステム利用環境や状況を知ることが、コンテキストウェアな対話戦略を備えた音声対話システムへの第一歩である。これまでも車につけられた多数のセンサ信号によって、運転者のストレス状況を推定する研究が行われている。安全性の尺度は、センサ信号の分析に基づく運転者の運転負荷としてリアルタイムで数値化する。

(2) 運転状況に応じた運転負荷によって、運転者の運転を妨げないような音声対話戦略を検討する。利用者の状況は時々刻々と変化する。例えば、「停止中」であれば比較的安全にシステムとの対話を行うことが可能だが、「走行中」であれば運転者はシステムとの対話をなるべく避けるべきである。例えば、運転者の運転負荷が低い場合には、複雑だが素早く検索できる対話戦略を採用し、運転者の

運転負荷が高い場合には、時間はかかるが単純な対話戦略をとる。

(3) 対話システムとのやりとりの背景には運転者の運転負荷も関わっている。すなわち、2において対話がスムーズに進んでいるか否かをリアルタイムで判定できれば、運転状況推定にもとづく安全性の推定がより精密になると考えられる。

## 3. 研究の方法

(1) 運転者の運転負荷を推定する方式を検討する。運転者の体や車両本体に様々なセンサをとりつけて、短時間の生体信号や運転情報信号を測定し、そこから運転負荷の推定を行う。一方で、これらのセンサは取り付けに専用の知識が必要で高価なことも多いため、入手が容易なスマートフォンで得られるセンサの利用を想定した運転負荷推定も行う。

(2) 運転中の利用に適した対話方式を提案するため、ユーザ負荷の観点で対話戦略を評価する。対話の主導権の方式と確認発話の方式を組み合わせ、計算機シミュレーションによる客観評価実験をおこなう。

(3) 対話システムとして適切に評価するためには主観評価実験が必要である。実際の車を運転した評価は危険であるため、シミュレーション実験による評価をおこなう。具体的には、運転動作と音声対話システムの利用という二重タスク課題における主観評価事件をおこなう。

(4) ユーザの負荷状況の推定結果を組み込んだ、柔軟な対話戦略の選択方式を実現するために、グラフ探索に基づく対話戦略の導入をおこなう。対話シミュレータによる客観評価実験により、その性能を評価する。また、ドライビングシミュレータを用いた主観評価実験によって、その性能を評価する。

## 4. 研究成果

(1) 生体信号と運転情報信号を収集したデータベース(NUDrive)を利用して、運転負荷の推定精度について評価を行った。NUDriveでは運転中に運転以外の様々なタスクを課している際のデータも含まれているため、タスクの有無を負荷の有無と仮定して実験を行った。

生体信号と運転情報信号を組み合わせた負荷の推定実験を行った。実験の結果、男女混合の推定器を用いた負荷推定は76%の精度であったが、女性のデータのみ対象として同様の推定を行ったところ81%の精度に向上した。従って、男女別に推定器をつくるのが有効であると考えられる。

続いて、有効な特徴量を検討するために個々の特徴量のみを用いた場合の推定精度を比較した(図2)。その結果、運転負荷の推定に最も有効な特徴量は加速度であることが示

された。

最後に、スマートフォンでの利用を想定して加速度と心拍数を組み合わせた特徴量による推定を行った(図3)。加速度のみで推定したほうが精度が高いという結果であった。すなわち、加速度センサのみを用いた運転負荷の推定で十分という知見が得られた。

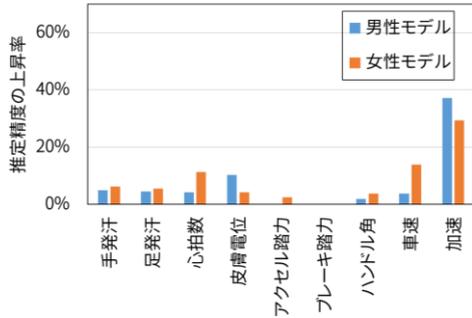


図2 特徴量ごとの推定精度の変化

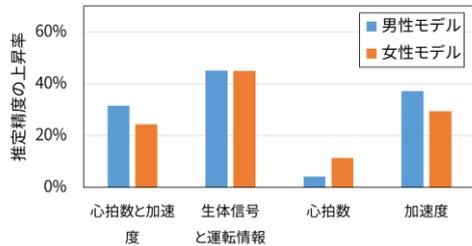


図3 スマートフォンで取得可能な信号に限定した場合の推定精度の変化

(2) 音声対話によるカーナビゲーションシステムは、運転中でも脇見運転をせず、安全に使用することができると期待されてきた。しかし、音声対話システムはユーザに認知負荷がかかり、運転に関わる反応が遅れてしまう恐れがある。その原因の1つは音声認識器の誤認識と、誤認識の回復を意図した確認対話と考えられる。特に、誤認識したまま対話を進めることを防ぐために、ユーザの入力ごとに確認発話を行うシステムが多いが、確認発話のために、タスク達成までのターン数や時間が増えてしまう。これは、ユーザが煩わしく感じてしまうという問題だけでなく、運転者の注意が運転行為からシステム操作に向かってしまい、交通事故の要因となる恐れもある。

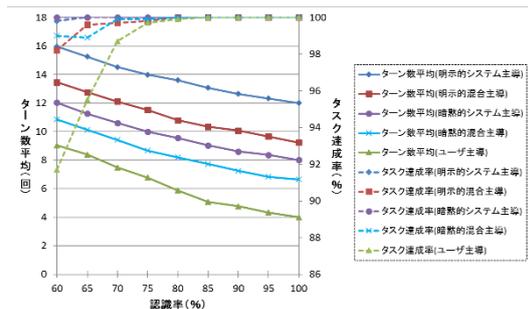
そこで、本研究ではユーザの負担を軽減するための対話戦略として「できるだけ早く応答するのではなく、余裕時のみ応答する対話戦略」について検討した(図4)。運転中に適した対話戦略を明らかにするために、対話の主導権の方式と確認発話の方式を組み合わせ、シミュレーション実験をおこなう。シミュレーション実験は、運転中の利用者に余裕がある場合とない場合を想定し、システムがユーザに余裕がある場合のみ発話する余裕時発話条件と、ユーザの肯定文と否定文の発話正答率を80%にするユーザ誤認識条件の2つの条件で行う。評価項目は、ターン数平均、タスク達成時間平均とタスク達成率である。



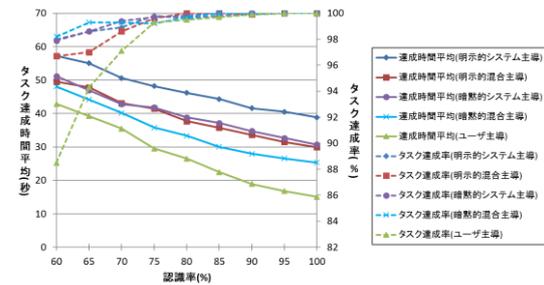
図4 余裕時のみ応答する対話戦略

まず、余裕時発話条件のシステムの認識率に対するターン数平均とタスク達成率の関係を図5aに、タスク達成時間とタスク達成率の関係を図5bに示す。図5aを見ると、ターン数平均はユーザ主導、混合主導、システム主導の順に増えていく。また、暗黙の確認発話より明示の確認発話の方が、ターン数が多くなる傾向にある。システムの認識率に対するターン数平均の傾きは対話戦略によらずほぼ一定であり、システムの認識率の変化による順序の交替も起こっていない。図3bを見ると、暗黙の確認発話を用いたシステム主導と明示の確認発話を用いた混合主導のタスク達成時間平均は同程度となっている。その他の対話戦略の順序はターン数平均と同じであることが分かる。

次に、ユーザ誤認識条件のシステムの認識率に対するターン数平均とタスク達成率の関係を図6aに、認識率に対するタスク達成時間平均とタスク達成率の関係を図6bに示す。図6aを見ると、ターン数平均の順序は余裕時発話条件と同じであることが分かる。図6bを見ると、タスク達成時間の順序はターン数平均と同じである。タスク達成率はシステムの認識率の低下に伴って低下してしまう。このことから、タスク達成率を一定以上にするためには、システムの認識率を上げることはもちろんであるが、ユーザの認識率を100%にする必要があるということがわかる。

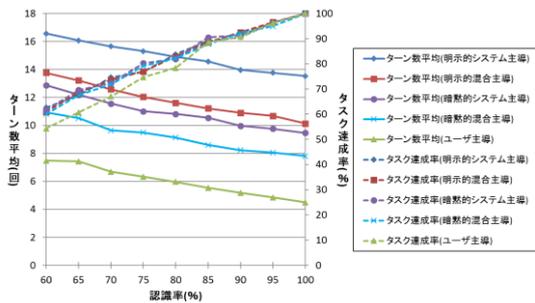


(a) ターン数と達成率

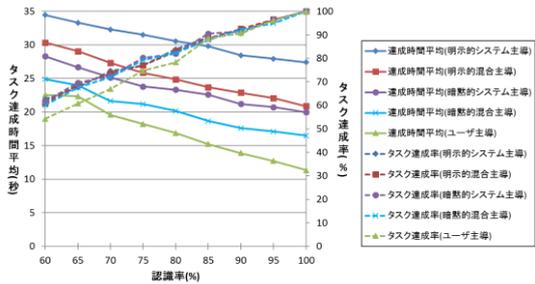


(b) 達成時間と達成率

図5 余裕時発話条件の実験結果



(a) ターン数と達成率



(b) 達成時間と達成率

図6 ユーザ誤認識条件の実験結果

また、余裕時発話条件(ユーザ認識率 100%)のシステム認識率 80%のときと、ユーザ誤認識条件(ユーザ認識率 80%)のシステム認識率 100%のときの、ターン数平均を比較すると、システム主導は同程度となる。一方、ユーザ主導と混合主導はユーザ誤認識条件の方が、ターン数平均は少ない。従って、ユーザの認識率が下がることよりもシステムの認識率が下がることの方が問題である。一般に、車内では雑音が多く、システムの認識率を 100%にすることは容易ではない。それに比べて、ユーザの認識率を 100%近くにすることは方法次第では可能である。例えば、余裕時発話条件のように、余裕がある場合のみシステムが発話することでユーザが対話に集中できるので、ユーザの認識率が上がり、ユーザ誤認識条件のようなタスク達成率の低下は起こりにくいと考えられる。これらのことから、余裕時発話条件はユーザの認識率を上げることができると考えられるため、タスク達成率を一定以上にするには有効であると考えられる。

(3) 自動車を運転する際、運転者には運転のための認知負荷がかかっており、運転中に音声対話を行うときは負荷がかかった状態で使用されると考えられる。本実験では、音声対話を行う第 1 課題を、ボール型ラジコンを操作する第 2 課題と同時にこなすという、2 重課題法で実験をおこなった。

走行コースは広い床に、底辺の一辺約 6 cm の直方体の障害物を一直線上に 50 cm 間隔で 5 本固定し、スタート地点とゴール地点はそれぞれ障害物の端から 50 cm 離れた場所とする。スタート地点からゴール地点まで、行きは障害物の間をジグザグ走行、帰りは並べられたペットボトルの横を直線で走行する。このコースを 1 回タスクが達成されるまで何周もおこなう。評価尺度は効率性、非負担度、満

足度とし、それぞれ 5 段階で評価をおこなってもらった。被験者は 4 名である。

想定単語認識率ごとの評価値を図 7a と図 7b に示す。3 つの評価値全てについて、認識率が 70%のときの評価値よりも 90%のときの評価値の方が高い。ユーザ主導は認識率 70%では全体的に評価が低く、特に非負担度が低い値となっているが、認識率が 90%になると評価値は高くなる。また、混合主導は認識率によらず、全ての評価値が高くなっている。これは、一発話目が何から入力してもよいという自由度の高さと、認識率の低下がターン数やタスク達成時間に影響を与えないという混合主導の特徴に起因していると考えられる。

余裕があるときのみ発話する場合は、非負担度は他の戦略と比べて最も高くなっており、音声対話が運転中にほとんど妨げにならなかったと言える。しかし、効率性や満足度が低いことから、システムの応答がすぐに返ってこないために、効率が悪いとユーザに感じさせてしまったと考えられる。システム主導において、認識率が 70%では、明示的確認発話よりも暗黙的確認発話の方が効率性と非負担度は高く、満足度は低い。これは、暗黙的確認発話の方が明示的確認発話よりもターン数は少ないが、聞き取るべき情報量が多く負担に感じたためだと考えられる。一方、認識率が 90%では、効率性は暗黙的確認発話の方が高いが、非負担度と満足度はほぼ同じ評価値となっている。これは、認識率が 90%だとシステムの誤認識が少なく、ユーザが明示的確認発話と暗黙的確認発話の違いをあまり感じなかったためだと考えられる。

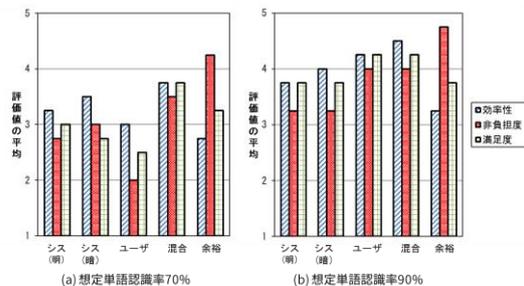


図7 主観評価実験の結果

(4) 車載用の音声対話システムにおいては、時々刻々と変化する運転者の負荷状況を考慮したシステム発話を行うことで運転者が注意散漫にならないような音声対話システムが望ましいと考えられる。タスク達成型の音声対話システムでは効率的にタスクを達成するための対話が目標であり、対話時間を短くするための対話戦略として、必要最小限の情報による簡潔で短いシステム発話を行うことが多い。ただし、簡潔な発話になるほど 1 語でも聞き逃すと文全体の意味が失われるか、誤って知覚され、運転者がシステムの発話を聞き逃してしまう可能性がある。聞き逃した場合、運転者はもう一度システム発話を聞かなければならないため、結果として対話時間が長く

なることも考えられる。したがって、運転中という状況下では必ずしも簡潔なシステム発話を行うのではなく、運転手に聞き逃されないようにやや冗長なシステム発話を行ったほうが、全体としての対話時間を短くできると考えられる。

運転手の行動という側面から考えると、運転手は聞き逃しがおこらないようにシステム発話を聞き取るという負荷を与えられている、と考えることができる。運転手は自動車の運転に注意を払うべきであり、対話システムの負荷は最小限に抑えることが望ましい。ただし、運転者の運転技術や運転経験によって運転に対する負荷は異なるため、二次タスクとしての音声対話にかけられる負荷の度合いは運転者によって異なるだろう。したがって、運転者の技量に合わせた対話戦略の選択が可能になれば、効率的かつ安全に配慮した車載用音声対話システムが実現できる。

そこで本研究では、これらの問題点を解決するため、運転者の運転操作や走行状況などから推定された負荷を入力することで、運転者の負荷が高いときには冗長な発話を選択できるような対話戦略（図8）を提案する。



図8 冗長な対話戦略

具体的には、システム発話選択の際に、発話の冗長性に関する尺度を新たに導入し、運転者の負荷を考慮することで、負荷が高いときに冗長な発話を選択されやすいようにする。提案手法の有効性を示すために、客観評価と主観評価を行う。客観評価では計算機シミュレーションを、主観評価では二重課題法をおこなう。

本研究では、グラフ探索アルゴリズムによる対話システムをベースとしてその対話戦略を拡張したシステムを実装した。グラフ探索アルゴリズムでは、対話によって得られた内容をスロットに埋めていく過程をグラフで表現する。グラフのノードはキーワード集合とし、理解候補の探索グラフを構成する。アクティブノードは現在の理解候補を表し、システムはユーザの発話から認識結果に基づいてノードを展開していく。また、グラフで複数のアクティブノードの存在を許すことによって複数の理解候補を保持することが可能である。複数の理解候補を表現できるため、選択中の候補が誤りであると気付いた際に別の候補を立てることによる回復が容易になる。

システム発話選択のための基準として無矛盾尺度  $S_c(q)$  と効率尺度  $S_e(q)$  に加え、システム発話の冗長性を考慮した冗長性尺度  $S_r(q)$  を導入した。

$$S_r(q) = R(q)$$

ここで、システム発話  $q$  が冗長な発話の場合は  $R(q) = 1$ 、簡潔な発話の場合は  $R(q) =$

0 である。ユーザの運転負荷が低いときは対話時間を短くするめ、簡潔な発話が望ましいが、ユーザの運転負荷が高いときはユーザが質問を聞き逃してしまう恐れがあるため、冗長な発話が望ましいと考えられる。

### ①客観評価実験

客観評価として計算機シミュレーションをおこなった。客観評価尺度はタスク達成時間、タスク達成率と肯定発話率を用いる。シミュレーションにはユーザシミュレータを導入して、運転負荷に対する敏感さを考慮した対話戦略の適切な選択方式について検討した。

ユーザシミュレータはランダムに目的地を決定し、システムの質問や確認に答えることによって、目的地の設定を完了しようと試みる。ユーザやシステムの話す速度は8モーラ/秒とし、発話と発話の間隔は1秒として計算する。本研究では、ユーザは負荷が高いときに話しかけられると、8割の確率でシステムの発話内容を聞き逃し、システムに再発話を要求するための発話を行うものとした。運転上級者と運転初心者の2つのユーザモデルとして、運転上級者は運転初心者に比べ、負荷が高い状態に陥っても負荷が低い状態に回復しやすいということを想定した。このモデルに基づくユーザシミュレータを客観評価実験に用いる。

運転上級者と運転初心者のタスク達成時間、タスク達成率と肯定発話率を図9と図10にそれぞれ示す。ユーザが聞き逃しをする場合の、冗長性尺度を用いた提案手法と冗長性尺度を用いない従来手法を比較する。どちらのユーザモデルにおいても、全ての評価尺度で、従来手法よりも提案手法の方が高い評価が得られている。また、運転上級者に比べ運転初心者の方が、従来手法と提案手法の差が大きい。これらの結果から、システムの発話内容を聞き逃しやすいユーザに対しては、提案手法は効果が高いことがわかる。

### ②主観評価実験

主観評価として二重課題法をおこなった。評価尺度は日本語版 NASA-TLX を用いる。簡潔な発話と冗長な発話を用いた音声対話システム

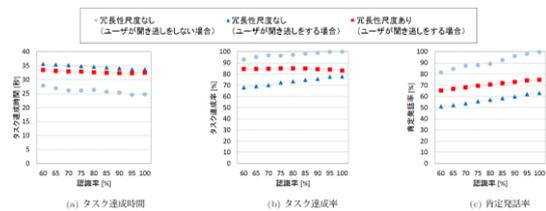


図9 客観評価結果（運転上級者モデル）

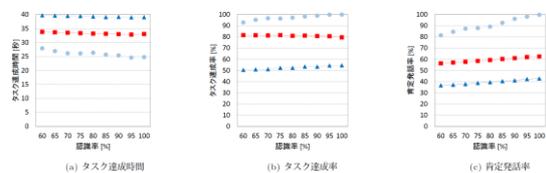


図10 客観評価結果（運転初級者モデル）

ムに対して、ユーザが感じる負担への影響を主観評価により検証する。

本実験では、第一課題としてドライビングシミュレータによる運転タスクを、第二課題として音声対話システムを使用することによる目的地検索タスクを行う二重課題法を行った。被験者は6名でそれぞれ運転習熟度が異なる。目的地検索タスクは、各対話ごとにあらかじめ指定された目的地を音声対話システムを用いて設定する、というものである。

音声認識には大語彙音声認識エンジン Julius を用いた。本実験では実車運転時の認識精度がそれほど高くはないという知見を意識して、平均認識率が80%程度となるような精度の低い音響モデルを用いた。

負担度の評価にはメンタルワークロードの主観的評価手法である日本語版 NASA-TLX を用いる。総合得点の算出には、AWWL (adaptive weighted workload) を用いる。AWWL は0から100までの値で、100に近いほどユーザが感じる負担が大きい。

被験者ごとの、簡潔な発話をする音声対話システムを使用したときの AWWL と冗長な発話をする音声対話システムを使用したときの AWWL の関係を図 11 と図 12 に示す。図 11 は直線コース、図 12 は市街地コースを運転した場合の結果である。被験者 C, D, E は簡潔な発話よりも冗長な発話の方が AWWL の値が高い。被験者 F は冗長な発話よりも簡潔な発話の方が AWWL の値が高い。この関係は走行コースに関わらず変化しない。つまり、運転に慣れている被験者は、音声対話システムを使用する際に感じる負担の評価が、運転負荷に影響を受けることなく安定していると言える。これに対して、被験者 A, B はコースによって AWWL が高くなる発話が異なる。これらのことから、被験者 A, B のような普段ほとんど運転をしていないユーザに対しては、運転負荷に応じて発話の冗長性を変化させる必要があると考えられる。

## 5. 主な発表論文等

[学会発表] (計6件)

- ① 鳥羽隼司, 原直, 阿部匡伸, “スマートフォンで収録した環境音データベースを用いた CNN による環境音分類,” 日本音響学会 2017 年春季研究発表会, 明治大学 (神奈川県・川崎市), 2017 年 3 月。
- ② 原直, 田中智康, 阿部匡伸, “重複音を含む環境音データベースにおける環境音検出のための特徴量の基本検討,” 日本音響学会 2016 年秋季研究発表会, 富山大学 (富山県・富山市), 2016 年 9 月。
- ③ 鳥羽隼司, 原直, 阿部匡伸, “RNN による実環境音データからのマルチ音響イベント検出,” 日本音響学会 2016 年秋季研究発表会, 富山大学 (富山県・富山市), 2016 年 9 月。
- ④ Masaki Yamaoka, Sunao Hara,

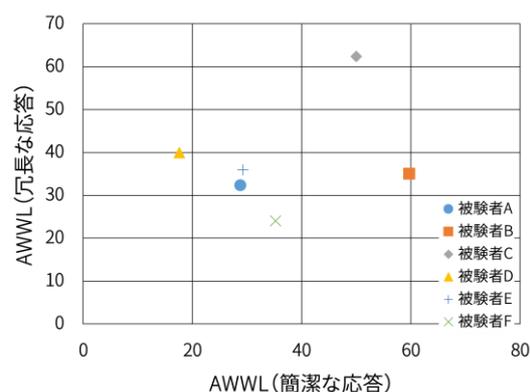


図 1 1 主観評価結果 (直線コース)

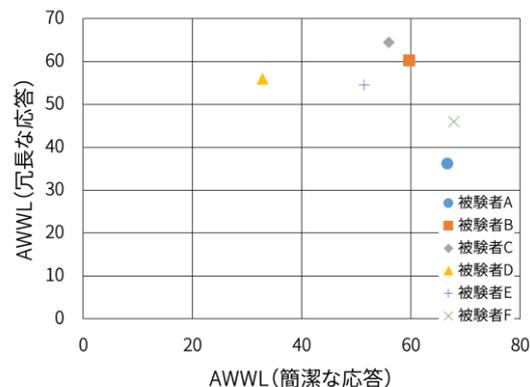


図 1 2 主観評価結果 (市街地コース)

Masanobu Abe, “A Spoken Dialog System with Redundant Response to Prevent User Misunderstanding,” APSIPA Annual Summit and Conference 2015, Hong Kong, 2015 年 12 月

- ⑤ 山岡将綺, 原直, 阿部匡伸, “冗長なシステム応答を用いたユーザの誤認識に頑健な音声対話システムに関する検討,” 日本音響学会 2015 年秋季研究発表会, 東京工業大学 (東京都・目黒区), 2015 年 9 月。
- ⑥ 山岡将綺, 原直, 阿部匡伸, “車載用音声対話システムにおけるユーザ負荷を考慮した対話戦略の検討,” 情報処理学会技術報告, 東京工業大学 (東京都・目黒区), 2014 年 5 月。

[その他]

ホームページ等 <http://www.a.cs.okayama-u.ac.jp/~hara/projects/kaken2014.html>

## 6. 研究組織

### (1) 研究代表者

原直 (HARA, Sunao)  
岡山大学・大学院自然科学研究科・助教  
研究者番号: 50402467

### (2) 研究協力者

阿部 匡伸 (ABE, Masanobu)  
山岡 将綺 (YAMAOKA, Masaki)  
鳥羽 隼士 (TOBA, Shunji)