Vol.2015-SLP-108 No.1
2015/10/30

車載情報機器の音声操作におけるタイムアウト制御の実験的評価

木下 悠[1]　川端 豪[1]

概要：筆者らはカーナビゲーションシステムに用いられているような階層メニュー構造を音声対話によってたどる操作を研究している．文脈にによる語彙の絞り込みは認識精度が向上することが考えられるが，以前にたどった文脈は無効化されているため言い直しが許容されない．本報では文脈とタイムアウト制御に基づいた語彙セットを有効化する方式について評価を行う．この方式を用いると，これまでたどった親の文脈がタイムアウトした後でもそれぞれが有効化したキーワードの再発話が可能となる．発話タイミングのモデルに基づく音声コマンドの連続発声をシミュレートし，単語誤り率とタスク達成率の観点から評価する．適切なタイムアウト時間はメニュー階層に関わることが示された．

キーワード：車載情報機器，音声操作，階層メニュー，タイムアウト

Evaluation of Timeout Control Methods for Voice Command Telematics Systems

Yu Kinoshita[1]　Takeshi Kawabata[1]

Abstract: We investigate the car operation system based on the tree-based item selection method via speech. Vocabulary reduction by command contexts is effective for increasing the recognition accuracy. However it does not permit the user to utter out-of-vocabulary words. This paper reports the new method to activate the word set based on the command context and its timeout. Using this mechanism, each keyword can be uttered afterward even when the parent context was terminated. The system is evaluated by WER and TCR measurements with simulated voice-command sequences generated by the speech timing model. Experimental results show that the appropriate timeout setting is related to the depth of the menu layer.

Keywords: Telematics, voice command, menu layer, timeout

1. 序論

これまで車載情報機器の操作は，視覚に基づく手操作を主体とするインタフェース (i/f) が主流であり，走行中の操作が脇見運転や前方不注意の原因となる危険性をはらんでいた．一方，近年のスマートフォンや情報家電などの発展に伴い，音声操作の利便性に注目が集まっている．この流れに沿って，もともとハンズフリー・アイズフリーが重要視される車載情報機器においても，音声対話 i/f の研究が活発化した．例えば，脇田らは運転中の情報機器操作の評価法を述べ，音声操作が視認手操作よりも利便性・安全性が高いことを定量的に示している [1]．また音声操作は手操作と比較して精神負荷が低いことも示されている [2]．このように車載情報機器における音声操作の有用性が研究される一方，それと同時に音声操作によって生じる負担やその安全性への影響にも目が向けられるようになった．音声操作をすることで注意が散漫になることや [3]，音声入力の回数が多いほど視認行動や運転精度が落ちることも示されている [4]．階層構造のメニューを音声操作する場合，階層を増やすと操作ミスが増加し，特に音声フィードバック

[1] 関西学院大学　理工学部
Kwansei Gakuin University,
School of Science and Technology

の待機時に注意が音声に行き運転が不安定になることがわかっている [5]．これらの研究から車載情報機器の音声操作について，適切な音声対話システムを検討していく必要性が明らかとなった．

清水らは利便性・安全性の観点から音声対話方式を検討し [6]，野田らは音声対話戦略の分析を行い，運転状況や負荷に応じて情報を提示する量を変える必要があることを示唆している [7]．また音声入力方法の観点から，「コマンド記憶」に由来する負担を軽減するため，より自然な文型に基づく発話による入力が考えられている [8][9]．このように車載情報機器の音声操作に関する様々な検討が進んでいるが，ユーザにとっての利便性・安全性をさらに高めるためには，車載情報機器における特殊性を考慮に入れる必要があり，筆者らはこれらに加え「妨害」を考慮に入れる．

車載情報機器の音声 i/f においては，物音や同乗者との会話による割り込みや，走行状況による危険回避などの運転操作による「妨害」を考えなければならない．割り込みの例として，運転手がシステムとの対話中に助手席の人が割り込んで発話した場合，システムとの対話を続けるか止めるかを判断する必要がある．また危険回避の例として，山道や急カーブなどの運転負荷が高い状況や，急な飛び出しが予想され，それに対する注意負荷が大きい状況では今までの対話を保存するか，始めからやり直すなどの動作を考えなければいけない．このように妨害後のシステムの振る舞いによって，i/f の操作性は大きく変わってくる．

前報においてこのような妨害のある状況下での音声 i/f の枠組みの検討した [10]．システムがあるキーワードを認識すると，それが文脈となり後続するキーワードを認識するための新たな語彙セットが有効化される．この語彙セットは次のキーワード認識の完了によって無効化されるのが一般的な文脈ベースコマンド処理の枠組みであるが，一度キーワードを誤認識すると既にその語彙セットが無効化されるているため，誤ったキーワードを再発話することができない．この問題に対応するために筆者らは各語彙に対する有効時間（タイムアウト時間）を設定し，文脈を遡ってキーワードを言い直すことができるような方式を考案した．

前報では上記の枠組みに加え，各語彙に対するタイムアウト時間を設定するための予備検討として，人間とシステムが対話して階層的なメニューをだどる実験を行い，メニューを順行，逆行する種々の場合でキーワードの発話間隔の傾向を調べた．この測定に基づく値をタイムアウト時間の標準値とし，分散による変動を観察した．

本研究ではシステムとの対話実験から得られた結果をタイムアウト時間の標準値とし，そこから値を調整してより厳しい設定から緩めた設定まで可変してその操作性を検討する．評価尺度としては，単語誤り率 (WER) とタスク達成率 (TCR) を用いる．

タイムアウトの考え方を導入する以前のシステムの動作をタイムアウト時間設定の特殊な場合として整理する．メニュー文脈によってある語彙セットが有効化され，そのキーワードの認識終了と同時に無効化する枠組みは，文脈に基づく語彙制限の方式として最も一般的なものと考えられるが，認識対象語彙のタイムアウトを導入したシステムにおいて，タイムアウト時間を 0 に設定することに相当する．この方式では文脈に基づいて認識対象語彙が最小数に絞り込まれるため，認識精度を高めるために有利である．しかし，文脈を逆行するキーワードの言い直しは許容されない．逆にメニュー文脈に関わりなく全てのキーワードを常に認識対象語彙として待機する枠組みのシステムは，タイムアウトを導入したシステムにおいて，タイムアウト時間を無限大に設定することに相当する．この設定では，文脈に沿わずにキーワードのスキップや言い直しが許容されるため発話の自由度が改良されるが，文脈に基づく認識対象語彙の絞り込み効果が期待できないため，認識精度の観点からは不利になる．

このように発話の自由度と認識精度はトレードオフの関係にある．本報では両極端のシステムをタイムアウト時間の設定という軸に基づき連続的に取り扱うことによって，トレードオフの調整を柔軟に行うことを試みる．すなわちタイムアウト時間を 0 から無限大の間で可変させてシステムの挙動を観察する．

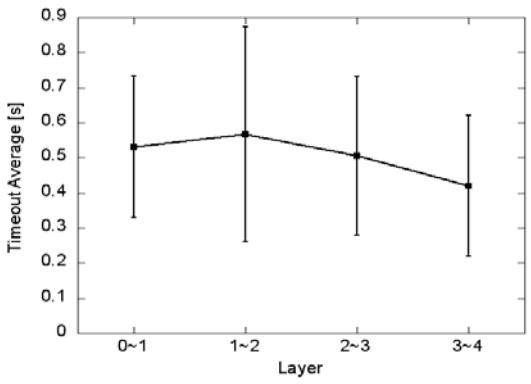

図 1 各階層における発話間隔

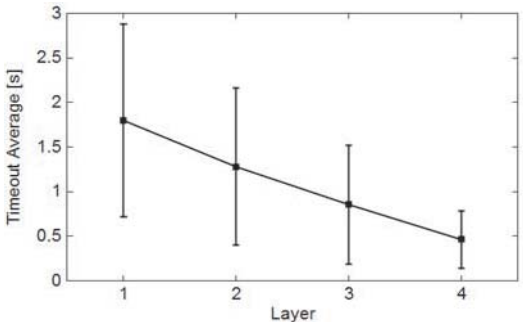

図 2 逆行先階層における発話間隔

2. タイムアウト制御を用いた音声操作

2.1 認識対象語彙のタイムアウト

階層的なメニュー構造を音声コマンドによって操作する際，ユーザが発声した単語が文脈となり，その文脈に対応した下の階層の語彙セットが有効化される．このとき発話中断が起きた場合，有効化されている文脈は一定時間保持され，この時間長をタイムアウト時間と呼ぶ．タイムアウト時間の設定方法にはいくつかの種類が考えられる．

タイムアウト時間を無限にする無限大型は，一度たどった文脈について再発話が可能であり，途中からの言い直しを考慮する場合には利便性が高い．しかしこれまでの全ての文脈に対する認識対象語彙が多くなるため，誤認識率が高まると考えられる．

また従来のように発声された文脈のみの認識対象語彙を有効化する場合は，有効化している語彙数が絞られるため誤認識が減少すると考えられる．これはタイムアウト時間が 0 になり，以前にたどった文脈についてはタイムアウトして再発話ができなくなる．このためより適切なタイムアウト時間の設定が必要である．

2.2 より詳細なタイムアウト設定方式の検討

前節において階層メニュー文脈における認識対象語彙のタイムアウト制御の基本的考えについて述べたが，具体的にシステムを実装するためには，より細部の動作を設ける必要がある．本報では複数の階層を遡って言い直しをできるように，ある時点で有効であった語彙は発話中のキーワード終了時点でタイムアウトを延長する．すなわち最後に発声された語彙の階層よりも上の階層の文脈のタイムアウト時間を更新する．このことで制限時間内では再発話が可能となり，また時間が経過するにつれて下の階層から順にタイムアウトしていくことになる．

それぞれの階層におけるタイムアウト時間の標準値は，前報においてより広い視点でもう一度測定範囲を拡大した結果を用いる．それらの結果を図 1，図 2（前項）に示す．図 1 は階層順行時における発話間隔を示しており，横軸は階層間，縦軸は発話間隔，ポイントは平均値，縦棒は標準偏差を示している．階層順行とは下の階層に順に進んでいくことで，この結果から順行時において階層の違いによる影響はみられない．階層逆行時とは再発話のことで，図 2 は最終階層から上階層への言い直しの発話間隔を示しており，横軸は逆行先の階層，縦軸は発話間隔，ポイントは平均値，縦棒は標準偏差を示している．逆行時には逆行先の階層が上に行くほど再発話による発話間隔は長くなっている．これは上階層ほど今までたどった文脈を思い出す量が多くなり，その文脈を思い出すことに時間がかかったためであると考えられる．階層逆行時の発話間隔の値は階層順行時と比較し長くなるため，システムに設定するタイムアウト時間は階層逆行時の発話間隔を参考にした．それらの値は上の階層から順に 4 秒，3 秒，3 秒，2 秒を標準値とした．

3. タイムアウト制御の実験的評価

3.1 実験方法

本研究の実験は，発話数が多くなり実際の被験者に行ってもらうには困難であるため，ユーザの発話シミュレーションによって評価データを生成した．音声には全てのシナリオを被験者 10 人に発声してもらい録音した音声を使用した．総シナリオ数は 59 個あり，1 個のシナリオは起動語を含め 5 個の単語が認識されれば，タスクが達成されるようなツリー構造になっている．具体的なメニューの階層構造の一部を図 3 に示す．起動語である「機器操作」を第 0 層とし，「機器操作」が発話されると第 1 層の語彙セットが有効化され，例えばタイムアウト時間内に「情報検索」と発話されると，第 2 層において「情報検索」が親である語彙セットが有効化される．このようにタイムアウトせずに第 4 層まで続けて認識されればタスク達成である．全 590 シナリオの中からランダムで 100 シナリオ選択，そしてそれらを連続的に再生した．また 3 回に 1 回の割合で再発話を行っている．発話者の発話間隔については，図 1 の観測に基づいて平均値を標準値とし，分散の幅だけ発話間隔をランダムに変化させ評価データを生成した．階層順行と逆行の場合の発話例を図 4 に示す．(a) は階層順行時におけるもので第 0 層から順に発話していく．(b) は階層逆行時で，この例においては第 0 層から発話を行い，第 3 層の「出発地」が発話された後，第 1 層の「ナビゲーション操作」に言い直しを行い，その後階層を順行していく発話である．一つのシナリオが終わると次のシナリオの始めから発話するようになっている．

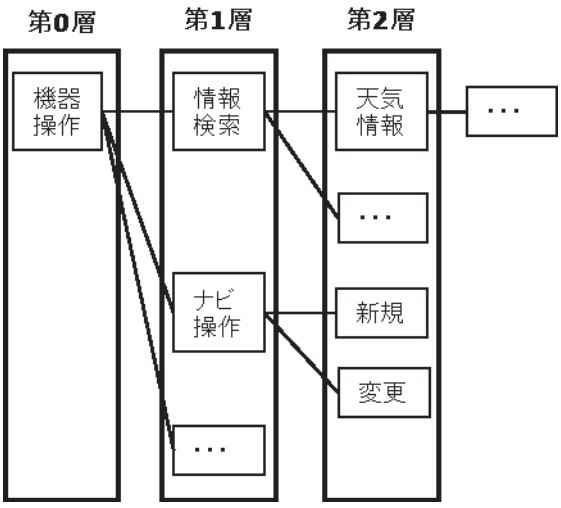

図 3 メニューの階層構造

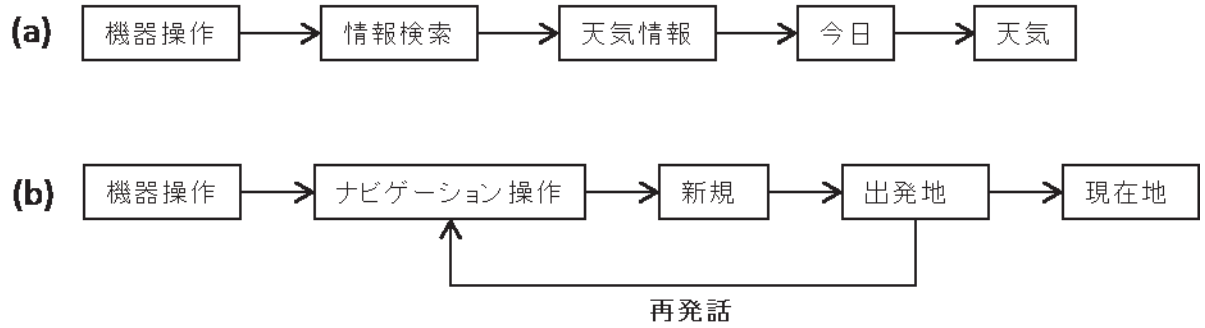

図 4 シミュレーション発話の例
(a) 順行 (b) 逆行

タイムアウト時間を設定値から-2 秒から+9 秒の範囲で補正し，また比較のためにタイムアウト無限大の実験を行った．評価項目は，タイムアウトの考えを導入して語彙を絞り込むことによる影響を観察するための単語誤り率 (WER) とシステム全体の評価をするために最終的な目的が達成されているかを確認するためのタスク達成率 (TCR) とした．WER は，S を誤認識の個数，D をタイムアウトしていて認識できなかった個数，I を雑音によって発話していな音が発話として認識された個数，N を発話総数として (1) 式のように計算した．またシナリオの 5 つの単語がタイムアウトすることなく認識されればタスク達成となり，TCR は，5 つの単語が誤認識されることなくまたタイムアウトすることなく認識され，タスクが達成された回数を T とし，タスクの総数を M として (2) 式のように計算した．

$$WER = \frac{S + D + I}{N} \tag{1}$$

S: 誤認識の個数
D: タイムアウトしていて認識できなかった個数
I: 発話していない音が発話として認識された個数
N: 発話総数

$$TCR = \frac{T}{M} \tag{2}$$

T: タスクの達成回数
M: タスク総数

3.2 実験結果と考察

音声認識を用いたシステムの評価において，基本となる音声認識精度の想定が必要になる．そこでまず録音した車内雑音と発話音声の SN 比と WER の関係を図 5 に示す．横軸は SN 比，縦軸は WER である．SN 比が 0dB から 3dB までは雑音が大きいため値が高くなっている．また 4dB から 8dB にかけて大きな変化はみられないが，4dB のときに値が最も低くなっている．これは雑音なしと比較すると，小さい雑音が入ることによって発話区間を検出する際の閾値が高くなったためであると考えられる．またこれから示す WER と TCR において，SN 比による傾向の違いは見られなかったため，これ以降は WER が一番低い SN 比が 4dB のときの結果を示し考察する．

図 6 にタイムアウト時間補正量と WER の関係を示す．横軸はタイムアウト時間の標準値からの補正量，縦軸は WER を示しており，標準値を用いたときの WER はタイムアウト時間補正量が 0 のときである．-1 秒，-2 秒では WER が高く，0 秒，+1 秒のときに最良値になり，その後は補正量が大きくなるにつれて高くなる．+7 秒から+9 秒においてはタイムアウト時間無限大と比較して差が無い．0 秒よりも低い値になるとタイムアウトした単語が誤認識として数えられるため WER が高くなったと考えられる．またタイムアウト時間補正量が増加するにつれて値が大きくなる要因は，認識対象語彙が多くなり誤認識率が増加したためであると考察できる．標準値付近では認識対象語彙が少なくなるため精度が低くなったと考えられる．このことからシステムにタイムアウトを導入し，標準値付近のタイムアウト時間を用いることによって認識精度が向上することがわかった．しかしタイムアウト時間が長くなるにつれて認識精度が落ちることから，適切なタイムアウト時間の設定が必要であると考えられる．

次にタイムアウト時間補正量と TCR の関係を図 7 に示す．横軸にタイムアウト時間の標準値からの補正量，縦

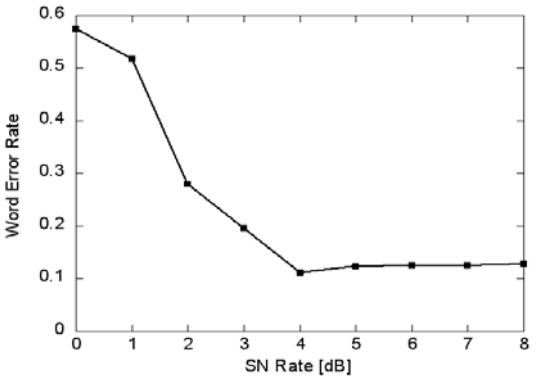

図 5 各 SN 比における WER

軸に TCR，ポイントは平均値，縦棒は標準偏差を示している．-1 秒，-2 秒では TCR が低くなり，標準値では高くなっている．これは WER の傾向と似ているが，TCR では+1 秒のときにより良い値になり，標準偏差をみると 0 秒と+1 秒の間に有意差があることがわかる．また+2 秒から長くなると徐々に値は下がる．-1 秒，-2 秒においてはタスクの途中でタイムアウトする語彙が多くなり値が低下したと考えられ，+2 秒以降で値が低下する要因は，認識対象語彙が多くなりタスクの途中で今たどっている文脈では有効化されていない語彙が誤認識され，不適切な文脈に移動してしまいタスクが達成できなくなるためである．このことからタスク達成率においても標準値付近で良くなることがわかった．

以上の結果から WER と TCR に基づく客観的評価によって，階層メニューの音声操作における標準値に若干の＋補正を加えたタイムアウト制御の有効性が確認できた．

4. 結論

車載情報機器の音声操作において，階層構造のメニューをたどる際にタイムアウト制御を導入し，単語誤り率（WER）とタスク達成率（TCR）の観点から評価を行った．実験ではより多くの発話が必要になることから，ユーザの発話シミュレーションによって評価データを作成した．

システムに導入するタイムアウト時間の設定法について，タイムアウト時間が 0 のときは認識精度は高いが再発話ができない．またタイムアウト時間が無限大では再発話が可能であるが，認識精度が落ちると考えられ，このトレードオフの関係についてタイムアウト時間の標準値に基づき値を可変させ評価を行った．

車内雑音と発話の SN 比によるシステムの挙動を観察したが，傾向に違いは見られなかった．そのため WER が最も低い SN 比の雑音を用いて実験を行った．発話間隔の傾向を調べることより得られる値を参考に標準値を決め，タイムアウト時間補正をその値の-2 秒から+9 秒と無限大で実験を行った結果，WER と TCR は標準値付近で精度が良くなった．これは語彙の絞り込みによって認識精度が向上したことを示し，またシステムの最終目的が達成される割合が大きくなったことから，システム全体としての評価が高いことを示している．どちらの評価項目も標準値よりも少ない値では，語彙がタイムアウトすることから急に精度が落ち，値が大きくなるにつれて徐々に精度が悪くなることがわかった．このため階層メニューを音声でたどる際にタイムアウト制御を導入し，標準値付近のタイムアウト時間を設定することで，音声 i/f の操作性が向上することが示唆された．

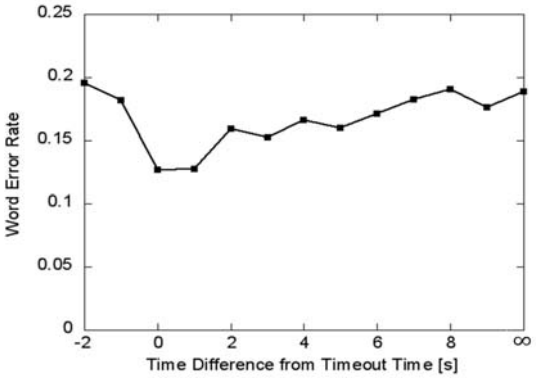

図 6　タイムアウト時間補正量別の WER

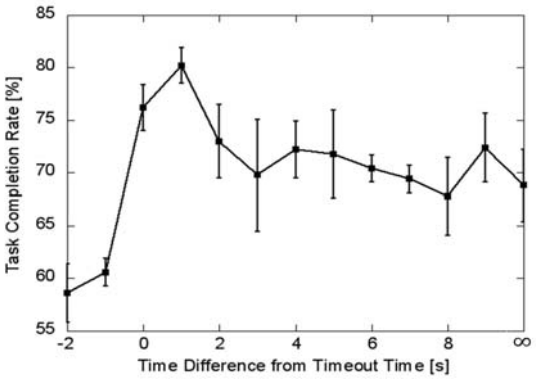

図 7　タイムアウト時間補正量別のタスク達成率

参考文献

[1] 脇田敏裕他：「運転中情報機器操作性の評価法」情報処理,Vol.42,No.7,pp.1762-1769 (2001)
[2] 伊藤一也他：「ドライビングシミュレータによる音声入力システムの評価」, 自動車技術会学術講演前刷集, No.93-03 (2003)
[3] John D. Lee 他：「Does a speech-based interface for an invehicle computer distract drivers?」World Congress on Intelligent Transport System (2000)
[4] 宇野宏他：「情報機器の操作が運転行動に与える影響に関する実験研究」, 自動車技術会論文集,Vol.45,No.2,pp.387-392 (2014)
[5] 沼田仲穂他：「カーナビ音声操作がドライビングシミュレータにおける運転パフォーマンスに与える影響」, 自動車技術会論文集,Vol.37,No.6,pp.181-186 (2006)
[6] 清水司他：「運転中における音声対話システムの評価」, 情処研報, 2000-SLP-64,pp.87-92 (2000)
[7] 野田幸志他：「心的負荷状況における車載情報機器のための音声対話戦略の分析」, 情処研報, 2006-SLP-64,pp.149-154 (2000)
[8] 倉田岳人他：「ユーザの発話傾向分析に基づく車載機器操作のための音声入力手法の提案」, 情処研報, 2009-SLP-78 (2000)
[9] Kenneth White 他：「Honda Next Generation Speech User Interface」, SAE paper, 2009-01-0518 (2009)
[10] 木下悠他：「車載情報機器の音声操作における文脈の保持とタイムアウト制御」, 情処研報, 2014-SLP-104 (2014)