E-040

語彙外発話に着目した音声インタフェースの評価

Evaluation of Speech Interface Watching Out-of-Vocabulary Utterance

瀧　直人[†1]・笹原 大亮[†2]・畑岡 信夫[†]　中野 鐵兵[‡]・小林 哲則[‡]

Naoto TAKI, Daisuke SASAHARA, Nobuo HATAOKA[†]　Teppei NAKANO, Tetsunori KOBAYASHI[‡]

あらまし　本稿では，音声インタフェースの課題と対処を明確にする目的で，音声のカーナビゲーションシステムへの応用を取り上げて，音声インタフェースの評価を実施した．その結果，語彙外発話が多発することと語彙外発話に気付くまでのユーザ挙動に関して新しい知見を得た．評価を実施したシステムは２つあり，一つは実際のカーナビ製品に具備されている音声インタフェース（認識）であり，他方は研究用として開発された音声認識ソフトウェア FlexibleShortcuts+ Select&Voice を使用した音声インタフェースである．

キーワード　音声インタフェース，音声認識，カーナビゲーションシステム，語彙外発話，FlexibleShortcuts，Select&Voice

１．まえがき

現在，車を所持している人のカーナビゲーション搭載率は約８０％であり，搭載８０％に対して約１０％の人たちが音声認識機能を使用している．しかし音声認識機能はうまく作動しない場合が多い．音声認識機能を使った人の代表的な意見は，「まずどう使って良いか分からない．」，「どのように発話して良いか分からない．」，「発話してもまるっきり別の言葉に認識され，使えない．」が大半で，「一度トライしたが，きちんと認識しないのでその後は使っていない．」である．他には利用者の声質や話し方，雑音や残響，語彙数によって，認識が劣化する，認識するまでに時間がかかる，手動で操作したほうが正確であるという意見も多い．現状では，無理して音声認識機能をカーナビに搭載しているという感じがある．商品の差別化をするためのオマケになっているのではないかという意見もある．しかし，車の中での音声インタフェースは，本来は，“hands free”，“eyes free”を実現するという安全・安心の観点から重要な要素技術である．オマケではなくカーナビゲーションの音声インタフェースは無くてはならない必須の機能になる可能性は十分にあると考えられている[1]．

本研究に関連する従来の音声インタフェース研究では，音声インタフェースの評価手法の問題を取り上げ，評価にあたっての定量的な尺度が必要であるとした研究成果がある[2]．さらに，音声インタフェースで特に問題になる語彙外発話を取り上げた研究としては，語彙制約のない認識(open vocabulary recognition)に関する研究が多く，英語での/ing/や/ation/などの頻繁に出現する部分単語列（graphone=subword）を既存の語彙単語に組み込んだネットワーク型単語列で表現する方式[3]や World Wide Web(WWW)に存在する語彙を適宜利用する方式[4]が提案されている．本研究では，カーナビゲーションにおける音声インタフェースの評価を実施して，語彙外発話の原因を解明して技術課題を明確にする．本研究で評価したシステムは２つあり，まず始めに既存のカーナビゲーション製品の音声インタフェース[5]である．次に，早稲田大学で製作した音声認識ソフトウェアである「FlexibleShortcuts+ Select&Voice」[6]を搭載したカーナビ応用システムを評価した．評価のタスクは，カーナビシステムで頻繁に用いられる行き先設定と POI (Point of Interest) 検索である．

第２番目の評価で使用する FlexibleShortcuts+ Select&Voice の操作は，音声入力とコントローラでの操作入力の両方ができる．特に，Select&Voice 機能は，発話語彙をカテゴリ毎に区分(select)して発話することで，認識語彙を狭めて結果として誤認識を避けるメリットを考えて考案された機能である．さらに，操作履歴と発話履歴がＰＣ内に記録される機能により，語彙外発話や誤認識原因を考察する事が出来る．評価内容は，語彙外発話に焦点をあて，政令都市地名を含んだ地名の検索を行うことを実施し，語彙外発話が起きた理由と，語彙外発話が起きた時の被験者の動作と行動を解析する．政令都市地名等の検索をタスクとして選んだ理由は，地名発話での区切り方に個人差があり，結果として語彙外発話が多発するタスクであるので選んだ．例えば，「仙台市太白区」等の「区」名は市名との一括発話が必要であるが，通常は「仙台市」と「太白区」とに分離して発話することが多く，発話カテゴリを設定してその窓（フレーム）での語彙を発話させる Select&Voice では語彙外発話が多発することになる．

２．音声インタフェース評価：既存製品

2.1 評価実験装置と評価タスク

本研究の第 1 段階の評価として，室内の静かな環境と実際に車に搭載したときの音声認識率の評価を行った．使用するカーナビゲーションは Carrozzeria AVIC-HRZ88G Ⅱ（図 2：パイオニア製）である．タスクは，オーディオ操作と行き先設定，及び POI 検索である．行き先設定は，

†東北工業大学工学部　(現在：[1](株)日立情報制御ソリューションズ、[2]クラリオン(株))
〒982-8577 仙台市太白区八木山 香澄町 35-1

‡早稲田大学理工学部
〒169-8555 東京都新宿区大久保 3-4-1

事前に与えた行き先を発声して設定するタスクと自由に行き先を設定する自由発話の 2 種類を行った．

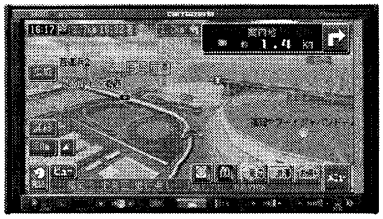

図 2　使用カーナビ（パイオニア製 Carrozzeria）

2.2 評価実験方法

まず始めに静かな環境での評価を行った．発話を正しく認識させるためには，次のことに気をつけて，話者に発声してもらい認識率を測定した．カーオーディオの音声を下げ，音声認識語を正しくはっきりと発話する．にごらない音をにごって発話すると，正しく認識されない原因になってしまう．また，読み方のルール(音声認識語)に従わない発話も正しく認識されないことがあるので注意してもらった．早口になったり，口ごもったりしないようにはっきりと明瞭に発話する．マイクは被験者の声を拾いやすい向きと距離に取り付ける．

評価実験では，振動防止のためにカーナビの左隣にコースターを敷き，その上にマイクを置いた．マイクの向きは被験者の方に向けた．

実験は男女合わせて２０人で行った．発話内容は①オーディオ操作，②こちらがあらかじめに用意した名称（キーワード発話），③被験者の好きな名称を発話（フリーワード発話），④住所と電話番号で自宅を検索の４通りとした．

次に実際に車に搭載して評価を行った．被験者は静かな環境で行った被験者から１０人を選び行った．実験内容は静かな環境（事務室）で行った内容と同じである．

2.3 評価結果

静かな環境と車載した時の音声認識率を比較した．結果を図 3 と図 4 に示す．図は発話の種類と何回発話して認識したかを表示している．オーディオ操作（コマンド入力）では，室内と車載での認識率を比べると，室内では 1 回で認識したのが 94%に対し，車載では 86%と 8%悪くなった．キーワード発話では，1 回で認識したのが，室内では 77%，車載では 64%と 13%悪くなった．この結果をみると，やはり静かな室内のほうが認識率が良いといえる．室内のフリーワードの語彙外発話 15%，車載のフリーワードの語彙外発話は 7%となった．この結果は，正式名称ではない，リストにないなどの語彙外発話と，同じ名称が多くて表示されないことが原因であった．キーワード発話 1 回認識では室内のほうが認識率は良いが，車載のフリーワードでは，車載の方が良い結果になっている．その理由は，被験者が実験を行う中で，使用方法を学習し，効率よく操作を行ったと類推される．

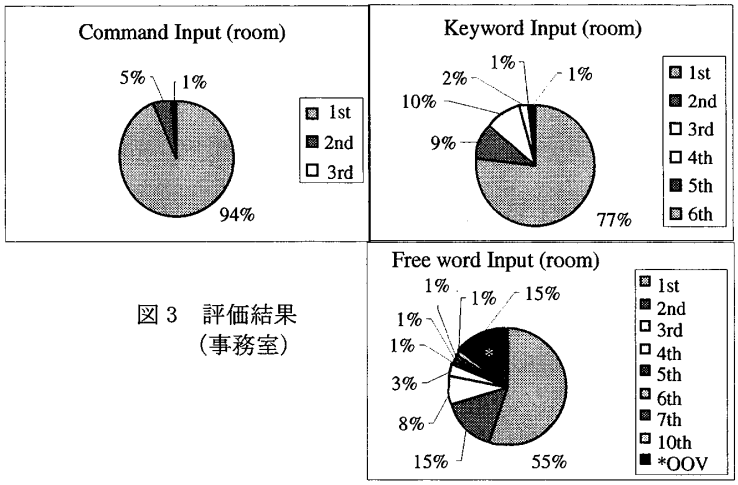

図 3　評価結果
（事務室）

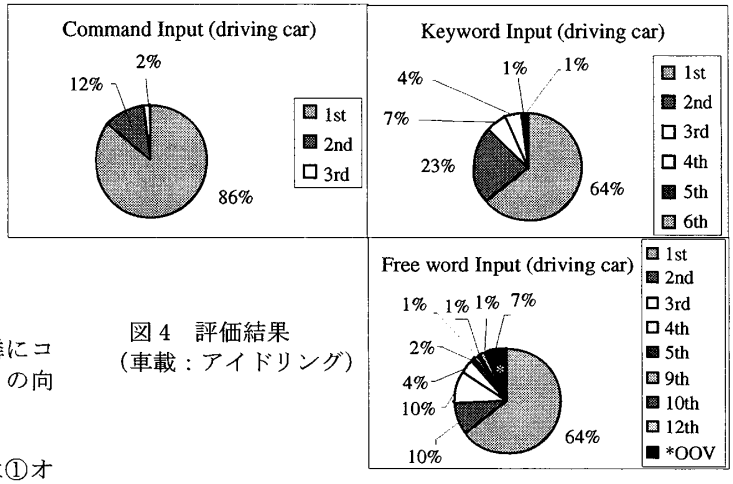

図 4　評価結果
（車載：アイドリング）

2.4 考察

評価を実施した結果，多くの問題が浮き上がってきた．話者の発話間違い，操作間違い，タイミング間違いは，話者が学習することによって改善できる．しかし，認識しない理由は他にも多くある．同じ名称が多くて表示されない（清水寺，本能寺など）．発話した言葉と場所が一致していても，正式名称でなければ認識されない（～カントリークラブを，～ゴルフクラブや～ゴルフ場と発話など）．略語に対応できない（マック，スタバなど）．語彙リストがない．音声リストがない（東京タワーなど，手動では検索できるが，音声認識では検索できないなど）．この問題を解決することができれば，さらに使い勝手が良いカーナビになるであろう．

3．音声インタフェース評価：プロトタイプ

3.1 FlexibleShortcuts と Select&Voice

音声インタフェースの評価第 2 弾として，早稲田大学にて開発された FlexibleShortcuts と Select&Voice を用いたシステムを使用した．早稲田大学では，経済産業省の委託を受けて，「音声認識基盤技術の開発」を進めており，その中で音声認識アプリ用共通プラットフォームの実現を目指して，Proxy-Agent の開発を行っている[10]．プラグインベースの機能拡張とサーバ連携の二つを特長としている．前者では，音声認識用資源の収集・管理・提供，

音声認識エンジンの逐次更新，推奨される部品利用法の共有等の機能を具備して，エンジン・アプリ非依存の枠組みを目指している．

FlexibleShortcuts と Select&Voice は，Proxy-Agent の枠組みの中で，アプリ開発用に作成された音声インタフェースである[6]．FlexibleShortcuts は効率的な機能選択用音声インタフェースであり，Select&Voice は，GUI とのアナロジーに基づいたデータ入力用音声インタフェースである．評価したシステムは，FlexibleShortcuts と Select&Voice を用いたカーナビのインタフェースであり，住所検索等の各種検索が可能なシステムとなっている．

3.2 評価実験の詳細

(1) 実験の目的

本実験では，FlexibleShortcuts + Select&Voice の枠組みで，ユーザが自分の語彙外発話に気が付くことができるかどうか，気付くとき，どのような過程で気付くことができるかどうかを調査する．意図的に語彙外発話が生じるよう実験を構成した上で，Proxy-Agent を通じて収集したログの解析により，ユーザの振る舞いを観察する．

(2) 収録環境

ログを正確に読み取り評価するため，雑音の入らない静かな大学の研究室である．

(3) 使用器具

早稲田大学で開発した車載向け PC アプリケーションを利用した．図 5 で示す PC とコントローラで構成され，

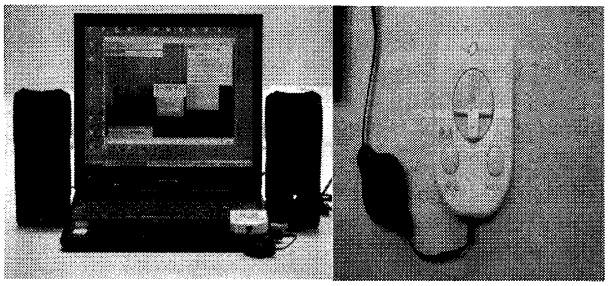

図 5　使用 PC とコントローラ

コントローラでのコマンド選択と音声認識でのコマンド選択のどちらでも入力が出来る．

(4) 被験者

操作経験有の被験者 5 名，経験無の被験者 5 名の計１０名とし，入出力のログと操作の画像を収録する．

(5) 発話条件

各被験者に計１０箇所の住所を検索させる．うち２箇所は市町村合併などで住所変更されていて，検索出来ない住所となっており，この２箇所の住所検索では語彙外が発生する．住所検索画面では，図 6 で示す様に，上から都道府県，市区町村，地域，番地と入力する場所（窓：フレーム）が決まっており，違う場所で発話すると語彙外発話となってしまう．

3.3 評価実験方法

まず，評価担当者が被験者に操作説明をし，被験者に簡単に操作してもらい本実験に移る．本実験はアドバイス無しで行い，実験中のＰＣ画面を評価の為に録画する．評価用１０箇所の住所を検索したら実験を終了し，操作ログ，発話ログ，録画映像にて評価を行う．

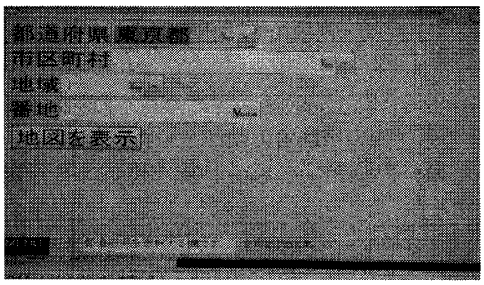

図 6　住所検索画面

3.4 評価結果

(1) 住所検索用評価データ

市町村合併などで住所が変わり，検索しても住所が出てこない住所を２箇所と，政令指定都市３箇所の住所，その他５箇所の住所の計１０箇所の住所を被験者に検索してもらう．下記に評価データの一部を示す．

埼玉県大里郡大里町相上 1-1
変更後：埼玉県熊谷市相上 1-1
茨城県鹿島郡神栖町息栖 1-1
変更後：茨城県神栖市息栖 1-1

(2) 語彙外発話の比率

表 1 に，操作経験有無での評価結果を示す．

操作経験の有る被験者の総発話数は 352 回で，その内 191 発話（54%）が 1 回で認識されたコマンドで，語彙外発話が 86 回（24%），誤認識が 74 回（21%），誤作動バグが 1 回だった．また，操作経験の無い被験者の場合は，総発話数 601 発話に対し，認識が 304 回（50%），語彙外発話が 151 回（26%），誤認識が 146 回（24%），誤作動バグが 1 回という結果だった．

表 1　操作経験有無での結果

	操作経験有		操作経験無	
	回数	割合	回数	割合
認識	191	54%	304	50%
語彙外	86	24%	151	26%
誤認識	74	21%	146	24%
バグ	1	0%	1	0%
計	352		601	

(3) 語彙外発話回数と種類

語彙外発話の種類は，住所検索タスクでは，①語彙外住所，②住所の区切り誤り（政令指定都市），③入力場所，④トップ画面 FlexibleShortcuts タスクでの語彙外発話の 4 種類があった．表 2 で示すように，全体の発話回数 953 回に対して，総語彙外発話は 237 回，その内検索出来ない語彙外住所は 112 回で語彙外発話内の 47%，住所の区切りミスは 55 回（23%），入力場所ミスは 21 回（9%），タスクでの語彙外発話は 49 回（21%）という結果になった．

表 2　語彙外発話

語彙外住所	住所の区切り	入力場所	トップ画面	計
112 回	55 回	21 回	49 回	237 回
47%	23%	9%	21%	

(4) 語彙外発話後の被験者の動作

語彙外発話の種類ごとに，語彙外発話後の被験者の動作を評価した．

(a) 語彙外住所

被験者に１０箇所の住所を検索してもらい，その中の２箇所に市町村合併などで住所変更されていて，検索出来ない住所を検索した時の被験者の行動を評価した．表 3 に市町村名が統合により新しい名称になっている住所検索で，4 回目の発声で気が付いた例を示した．さらに図 7 に語彙外発話であると気付くまでの回数を示した．検索出来ない住所だと気付くまでに個人差はあるが，1 回目が 0 回，2 回目が 1 回（5%），3 回目が 3 回（15%），4 回目が 4 回（20%），5 回目が 1 回（5%），6 回目が 2 回（10%），7 回目が 6 回（30%），8 回目が 0 回，9 回目が 3 回（15%）という結果になった．

表 3　語彙外住所の例

語彙外住所の例	**（4 回目で気付いた例）**	
発話内容	**動作内容**	**誤認識の理由**
茨城県	茨城県	
鹿島郡神栖町	かすみがうら市	語彙外
鹿島郡神栖町	猿島郡五霞町	語彙外
鹿島郡神栖町	猿島郡境町	語彙外
鹿島郡神栖町	猿島郡境町	語彙外
次の住所検索へ		

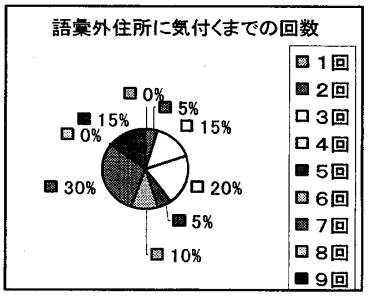

図 7　語彙外発話に気付くまでの回数

(b) 住所の区切り位置

政令指定都市を検索する際，市区町村の欄で，～市～区と発話しなければならないが，市名だけしか発話しない事が多い（例.札幌市中央区を札幌市と発話）．また，検索する住所を区切らずに読むと認識されない（例.「北海道札幌市中央区北三条西」と発話し，「北海道」と認識）．この結果，被験者は住所の区切り位置を理解し，「札幌市中央区」と発話できる場合が多々あった．

(c) 入力場所

住所検索をする際，カーソルを都道府県に合わせたまま，次の市区町村を発話してしまうと認識されない（例.カーソルが都道府県のまま「仙台市太白区」と発話し「山梨県」と認識）．この結果，被験者は山梨県を宮城県に直し，カーソルを市区町村に合わせ，「仙台市太白区」と発話できた．

3.5 考察

住所検索での新規住所の検索では，旧住所名を発声し，結果として語彙外発話となることが多かった．語彙の規模を気にしない場合は，旧住所もすべて語彙として登録しておけば良いが，その場合は認識率は劣化することになり，対処策が難しい問題である．今回の実験では，新住所名に変更されていると大半の人が気付くには 7 回程度の発話が必要であることが分かった．実製品では，システムからの注意等を与える等の配慮が必須であろう．

4．まとめ

本論文では，カーナビゲーション応用での音声インタフェースの評価を，語彙外発話に着目して実施した．実際の製品での音声インタフェース評価の結果，語彙外発話が多発することを明確にした．次に，第 2 段目の FlexibleShortcuts+ Select&Voice を使用した試作システムでの評価では，語彙外発話が起こったときの使用者の挙動を解析した．語彙外発話が発生すると被験者も学習をしていき，同じ間違いをする事は少なくなっていた．

謝辞

経済産業省委託研究「音声認識基盤技術の開発」の研究支援を頂いて実施した．関係各位の皆様に感謝する．

文 献

[1] 早稲田大学 IT 研究機構　音声技術実用化研究所「音声認識技術実用化に向けた先導研究成果報告書」A－83～B－6（平成 18 年 3 月）．

[2] 石川泰他：音声インタフェースの評価，日本音響学会会誌 61 巻 2 号，pp. 79-84（2005 年）．

[3] K.Vertanen, "*Combining Open Vocabulary Recognition and Word Confusion Networks*," ICASSP2008, pp.4325-4328, Las Vegas (Apr. 2008).

[4] S. Oger, et al., "*On-Demand New Word Learning Using World Wide Web*," ICASSP2008, pp4305-4308., Las Vegas (Apr. 2008).

[5] N.Hataoka, et al., "*Evaluation of Interface and In-Car Speech – Many Undesirable Utterances and Sever Noisy Speech on Car Navigation Application-*," Proc. of MMSP2008 (Sep. 2008)

[6] T. Nakano, S. Fujii and T. Kobayashi, "*Extensible Speech Recognition System using Proxy-Agent*, " Proc. of ASRU2007, pp.601-606, (Dec. 2007).

[7] 瀧直人，笹原大亮，畑岡信夫，中野鐵兵，熊井 朋之，小林哲則：カーナビにおける音声インタフェースの評価－ 語彙外発話の状況と対応案に関して －，信学技報，Vol. 108, No. 465, SP2008-152（2009 年 3 月）

[8] N. Hataoka, et al., "*Robust Speech Dialog Interface for Car Telematics Service*," Proc. of IEEE CCNC2004 (Jan. 2004).

[9] Y.Obuchi, et al. "*Development and Evaluation of Speech Database in Automotive Environments for Practical Speech Recognition Systems*," Proc. of Interspeech2006, Pittsburgh, PA, USA (Sept. 2006).

[10] 中野鐵兵，佐々木浩，藤江直也，小林哲則：「集合知を利用した語彙情報の収集・共有・管理システム」，情処学 SLP 予稿集，Vol2008, No.96, pp.77-84（平成 20 年 5 月）．