言語モデルを用いたカーナビゲーション音声認識技術の開発

Development of Japanese Speech Recognition Technology using Language Model for Car Navigation system

川添　佳洋，小林　載，吉田　実，外山　聡一
Yoshihiro Kawazoe, Hajime Kobayashi, Minoru Yoshida, Soichi Toyama

要　旨　多様なユーザー発話への対応と初心者でも使いやすい音声認識インターフェースの実現を目的に、言語モデルを用いた音声認識技術を開発した。

カーナビ操作に関連した様々なテキストデータを独自に収集して、カーナビ操作に適した言語モデルを作成するとともに、カーナビに搭載することを視野に入れて、認識処理の効率化を図った。

音声認識実験により、不要語を含んだ発話のキーワード認識率が大幅に向上することを確認した。

Summary　We developed a Japanese speech recognition technology using language model for car navigation systems, anticipated to realize an intelligible speech recognition interface for unaccustomed users and to cope with versatile speech styles.

Various text data related to the operations of the car navigation system was collected uniquely by the team, as well as the creation of an appropriate language model, leading to progress in the recognition processing efficiency aimed at car navigation systems.

Experimental results show that the key-words recognition performance has improved on speech including redundant utterances.

キーワード　：音声認識，言語モデル，カーナビゲーション

1. まえがき

当社の国内市販カーナビゲーションは音声認識機能が搭載されており，走行中でも音声を使って安全に操作が可能になっている。2003年発売のカーナビゲーションからワードスポッティング機能が搭載されている。これは，カーナビ操作コマンドの前後に不要な言葉を付与して「えーっと，100メータースケールにして」と発話してもカーナビの操作に必要な「100メータースケール」が認識できる技術である。この機能によって，ユーザが思わず口にしてしまった「えーと」「あー」などの言葉を含んだコマンドでも誤動作するケースが減少した。

今回，さらに発話の自由度を向上させ，初心者でも使いやすい音声インターフェースを実現するために，言語モデルを用いた連続音声認識技術を開発した。

この認識技術はＰＣ向けとして製品化されている音声認識ソフトでも用いられており，これらのソフトは，入力発話を全て認識してテキスト化する口述筆記が可能となっている。さまざまな分野の口述筆記を実現するには数万から数十万語の大語彙が対象になるため，高性能なＣＰＵ，大容量のメモリが必要となる。したがって，これらの多くはＰＣやサーバー・クライアントシステム向けに提供されることが多く，組込み機器向けに製品化されているものはまだ少ない。

我々はこの認識手法をカーナビの音声操作に用いることにした。まず始めにユーザーの発話を全て認識した後に，カーナビ操作に関連するキーワードを抽出することによって，「200メータースケール」と操作

コマンドのみを発話した場合の認識性能を劣化させることなく，長い不要語を含んだ「えーとそれじゃー地図の表示を 200 メータースケールに変えて」などと発話しても認識し，操作が可能になると考えた。このようにユーザーの多様な発話に対する認識性能が向上する反面，先に述べたように演算量の増加は避けられない。しかしながら，カーナビの操作コマンドの認識は口述筆記の場合と異なり，語彙数が数千語規模の認識であること，また，発話の中から操作に必要なキーワードが正しく認識できれば良いことから，処理を工夫することによって演算量の増加を最小限に留めることが出来ると考え検討を行った。

また，言語モデルについても上記と同じ理由で，小規模のテキストデータでも十分な性能が得られることから，独自にカーナビ操作に関連するテキストデータを収集し，言語モデルを作成した。

そして，キーワード抽出については高度な言語理解技術を使うことなく，極めて単純であるが効果の高い方法で実現した。

以下，開発した音声認識エンジンのブロック構成，言語モデルの作成方法，キーワードの抽出方法について説明し，最後に性能評価の結果について報告する。

2. ブロック構成

言語モデルを用いた連続音声認識技術はさまざまな方法が研究されており，それらについては参考文献 (1) に詳しく説明されているので，そちらを参照されたい。

現在は 2 パス方式が主流となっているが，これは入力音声に対して音声認識を 2 回行う方法である。第 1 パスで，ある程度の精度の音響モデルと言語モデルを使って粗い認識結果を得た後，第 2 パスで，第 1 パスの中間結果を元に高精度の音響モデルと言語モデルで結果を確定する手法である。最初から高精度な音響モデルと言語モデルを使い，1 回の認識処理で結果を確定するよりも効率的な探索が行えることが知られている。

今回の開発では上記の 2 パス方式を用いることにした。この方式を使う利点は第 1 パスの仕組みは，既存の音声認識エンジンの処理を流用することにより実装が非常に容易に行えることである。しかし，一般的な手法をそのまま用いると計算量が増加してしまい，カーナビのような組込み機器でのパフォーマンス悪化が懸念される。そこで認識性能を落とすことなく，処理量の削減を検討した結果，**図 1** のようなブロック構成とした。現在は発話終了後に行う必要のある第 2 パスの処理量を削減していることが特徴となっている。さらに第 1 パス，第 2 パスで同じ音響モデルを使うことでデータ量を増加させることなく，高精度な連続認識を実現した。以下, 各ブロックについて説明する。

2.1 音響モデル

製品への搭載を考慮した場合，音響モデルの変更は影響が大きい。現在使用している音響モデルは，大語彙認識や固定文法を用いた連続認識の認識性能が高く，それを用いることで十分な性能が得られると考えた。そこで, 今回は音響モデルの変更は行っていない。

2.2 言語モデル

口述筆記を行う場合は，音響モデルで音声の類似性を求めるだけでは高精度な連続認識を実現することは難しい。そのため，単語 N グラムモデルに代表される統計的言語モデルを併用し，文法としてどの位出現しやすいかも含めて処理を行うのが一般的である。単語 N グラムモデルは文の先頭にはどのような単語が表れやすいか，また，ある単語の後ろにはどのような単語が接続しやすいかということを統計的に求めたものである。ここで，N は接続を考慮する単語数で，1 パス認識に単語 2 グラム (バイグラム)，2 パス認識に単語 3 グラム (トライグラム) を用いることが多いようであるが，最近の研究では単語 3 グラム以上の高精度モデルを使う例も見受けられる。

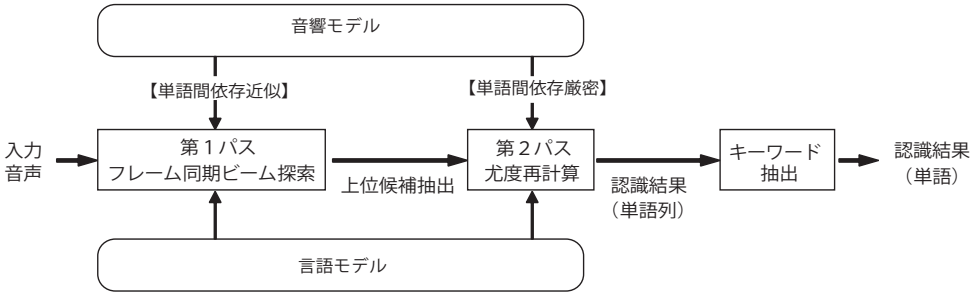

図 1　音声認識エンジンのブロック構成

言語モデルのパラメータを推定することを学習と言うが，それには大規模なテキストデータが必要となる。加えて，音声認識対象のタスクに一致するテキストデータで学習した言語モデルを用いた方が高い認識率が得られる。ＰＣ向けのソフトのようなさまざまな分野の口述筆記を行うことが要求される場合には，数十ヶ月分の大量の新聞記事を使って言語モデルを学習することによって，データ量とタスクの多様性の課題に対処している。

本開発の場合，「カーナビゲーションの操作」という一つのタスクを扱うため，多くのテキストデータは必要ではないと考えられる。その一方で，カーナビ操作独特のコマンドが数多く存在する極めて特殊なタスクであると言える，そのため新聞記事のような一般的な事柄を扱うテキストを使っても十分な認識性能が得られないという懸念がある。つまり，カーナビ操作に特化した言語モデルを作成するためには，それに関連したテキストデータが必要となる。しかし，このようなテキストデータを大量に入手することは非常に困難であるため，今回は独自に収集することにした。その結果，約１万文のテキストデータを集めることができた。学習データ量としては決して多くはないが，後述の認識実験で示すように，発話中のキーワードの抽出に関しては高い性能が得られている。

2.3　第１パス

音声入力と並行して認識計算を行い，発声が終了したと同時に認識結果を出すことが可能なフレーム同期型の認識計算を行っている。また，効率良く認識処理を行うために，処理の途中で認識候補の絞込みを行っており，計算の途中で入力音声との類似度が低い候補は計算処理を打ち切ることで計算量の削減を実現している。

各フレームで最終状態に達した単語の情報を保存しておき，第１パスの処理が終了した後に単語グラフからＮベスト候補を抽出し第２パスで再評価を行う。

単語グラフは音声認識では認識単語列の表現方法の一つであり，**図２**のように表される。始端Ｓから終端Ｅを結ぶ１つの経路が１つの認識結果に相当する。このように経路をたどることによって多くのＮベスト候補を得られるため効率の良い表現方法と言える。

2.4　第２パス

第２パスの処理は発話終了後に行っているため，このブロックの処理時間がそのまま結果出力の待ち時間に加わることになる。したがって，最小限の処理で最大限の効果が得られるような構成にすることが重要である。検討の結果，第１パスで得られた上位候補について，単語間の音響モデルを最適なモデルに差し替えて再計算を行うだけに留めた。

2.5　キーワード抽出

第２パスで得られた単語列からナビの操作コマンドに相当するキーワードを抽出する。高度な言語処理技術を使って意味理解を行う方法も考えられるが，今回は「認識単語列の最後に出現した操作コマンドを結果として出力する」という非常に単純なルールを用いた。このようなルールにした理由は，ユーザが複数の操作コマンドを発話する場合，「ノーマルビュー，じゃなくてスカイビュー」「えー，100 メータースケール，あっ間違えた，50 メータースケール」というようにユーザが希望するコマンドが最後に現れる場合が多いためである。このような単純な仕組みでもカーナビの操作が目的であるので，発話の意味まで理解することはできなくても，これで十分な効果を得ることができると考えられる。

なお，認識結果の単語列にキーワードが全く含まれていない場合は候補が無いと判断してリジェクトする仕組みになっている。

3.　認識実験

一般的に連続認識の評価は単語正解率と単語正解精度を使って，単語列がどの程度正解しているか，また挿入誤りや削除誤りは何語くらいかを評価する。しかし，カーナビ操作の場合は操作コマンド，つまりキー

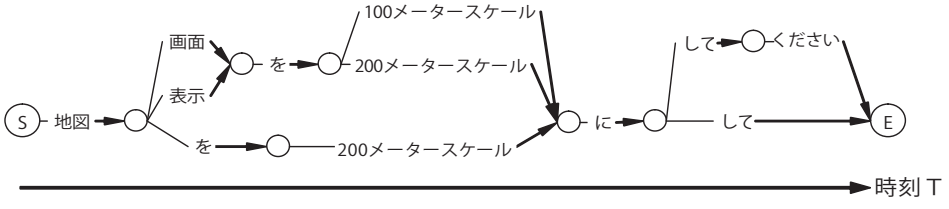

図２　単語グラフ

ワードが正しく認識できていることが重要であるので，ここではキーワード認識率で評価を行うことにした。不要語なし発話と不要語あり発話を用意し，シミュレーション実験によりキーワード認識率を調べた。

ここで，不要語なし発話とは，カーナビの操作コマンドのみを発話した音声，不要語あり発話とはカーナビの操作コマンドの前後に「えーと」や言回しを付けて発話した音声のことを言う。発話例を**表 1** に示す。

それぞれ，停止状態の自動車内で収録し，それに走行騒音を重畳した 2 種類の音声データで評価を行った。

不要語なし発話の認識結果を**図 3** に，不要語あり発話の認識結果を**図 4** に示す。比較として単語認識の結果を併せて載せた。

不要語なし発話の場合，どちらの認識方法も停止状態，走行状態に関わらず高い認識性能が得られている。若干，単語認識よりも開発手法の方が劣化しているが，ごく僅かであり実使用上は問題ないと思われる。

一方，不要語あり発話の場合，単語認識のキーワード認識率は不要語の影響受け大きく低下している。それに対して，開発手法のキーワード認識率は高い値を維持している。走行状態でも 80％以上のキーワード認識率が得られており，開発した手法の効果が非常に高いことがわかる。

4. まとめ

カーナビゲーション向けの言語モデルを用いた音声認識技術について報告した。従来の音声認識エンジンの仕組みを生かしながら，効率良く 2 パス処理を行う方法を検討した。さらに，カーナビ操作に特化した言語モデルを使うこと，単純ではあるが効果の高いキーワード抽出を行うことによって不要語を含んだキーワード認識率が従来に比べ大幅に向上した。これ

表 1　評価用音声データの発話例

発話例	
不要語なし	不要語あり
現在地	現在地を表示する
ノーマルビュー	えーとノーマルビュー
500メータースケール	500メータースケールに変えて
自宅	えーとじゃあ自宅までのルートを引いて
ヘディングアップ	画面をヘディングアップに切り換えて

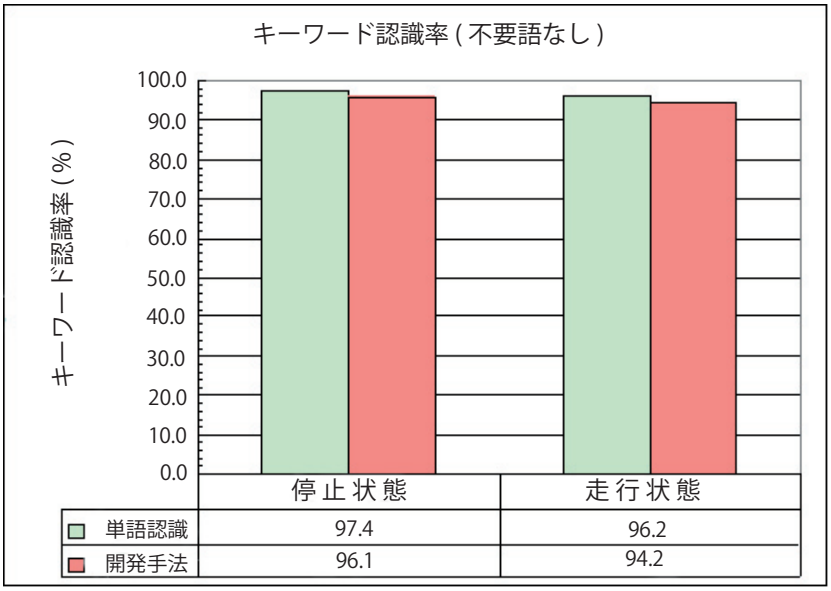

図 3　キーワード認識率（不要語なし発話）

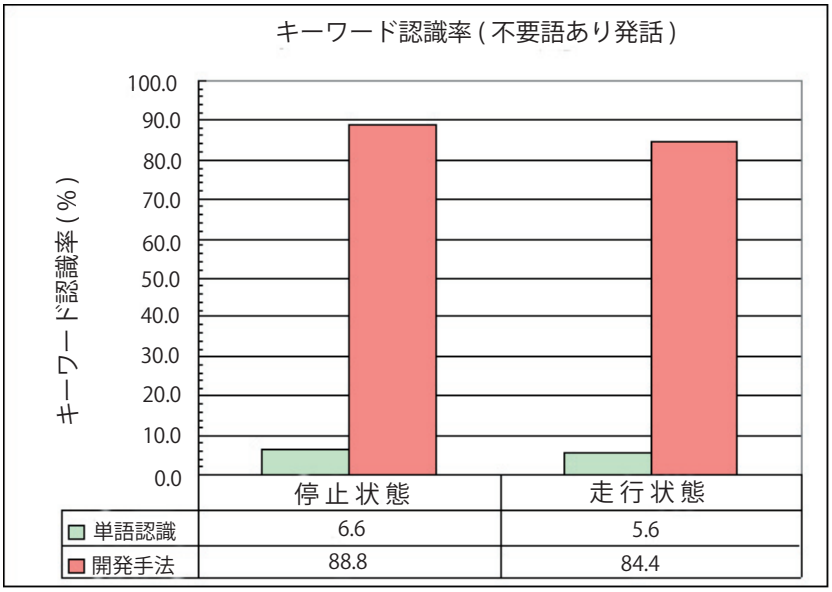

図 4　キーワード認識率（不要語あり発話）

により，多様なユーザ発話に対して高い認識性能が得られ，より使い易い音声インターフェースが提供できると考えられる。

今後はさらなる認識性能の向上，大語彙化などについても検討していきたい。

参考文献

(1) 鹿野清弘　他, "IT Text 音声認識システム ", オーム社, 2001
(2) 北研二, " 確率的言語モデル ", 東京大学出版会, 1999
(3) 河原達也, 荒木雅弘, " 音声対話システム ", オーム社, 2006

筆 者 紹 介

川 添　佳 洋（かわぞえ　よしひろ）
技術開発本部開発センターＭＳ第一開発部
小 林　載（こばやし　はじめ）
技術開発本部開発センターＭＳ第一開発部
吉 田　実（よしだ　みのる）
技術開発本部開発センターＭＳ第一開発部
外 山　聡 一（とやま　そういち）
技術開発本部開発センターＭＳ第一開発部