E-044

車載音声の解析と評価

Analysis and Evaluation of In-Vehicle Speech

上野 聡　齋藤 康夫　高橋 英徳　畑岡 信夫（東北工業大学）
Satoshi Ueno　Yasuo saito　Hidenori Taikahashi　Nobuo Hataoka

1.はじめに

今日、車社会でカーナビ搭載率は非常に高い。しかし、音声認識機能を使っている人は少ない。その理由として、音声認識技術に問題があるといえる。音声認識技術は、入力された音声を文字に変換し、規定された単語のどれかを判定する技術である。現状の問題として、

・実環境で動作しない(車載雑音、発話者以外の声で誤動作等)
・使用方法が分かりづらい(何を言っていいか分からない、話し始めのタイミング)
・語彙外発話の際正しく認識されない
・誤認識が生じると、システムの動作を利用者が理解できなくなることがある

等がある。その結果、使い方が難しく性能の悪いスイッチでしかないことが想定される。本研究では車載音声を取り上げ種々の周囲雑音の除去を行い、認識率の向上を図った。

2. 音声認識実験の概要

2. 1 大語彙連続音声認識ソフトウェア Julian

今回認識実験を行うにあたり Julian を使用した[1]。Julianは、有限状態文法(FSG)に基づく連続音声認識パーザである。Julian は言語制約以外のほとんどの部分を Julius と共有している。Juliusはn-gramで次の単語の予測を行っていくが、Julianはあらかじめ文法を設定する必要がある。今回は小語彙単語認識の枠組みでJulianを使用する。

2. 2 実験データ

今回使用したデータは、日立製作所と早稲田大学から入手したものを使用した[2]。

この音声データは、以下の条件のもとカーナビが使用される環境で収録された。

(1)自動車内に設置された遠隔マイクで、実際に走行中の音声
(2)単一マイクではなく、マイクロフォンアレイ
(3)孤立単語音声認識を対象
(4)読み上げ音声ではなく、自然発話に近い音声
(5)音声認識操作失敗時データをそのまま残す
(6)安全上、運転者ではなく助手席の声をそのまま残す。

(1)について、周囲雑音なども含めた環境の再現を意味する。走行時の音声認識性能を左右する因子としては、100Hz 以下に存在する車のエンジン音、窓の開閉、オーディオやエアコンのオンオフ、雨音やワイパーの影響、風切り音やタイヤノイズに影響する走行速度や路面状況、エンジンの回転数などが挙げられる。本研究では、あくまでも現実の状況を再現するという視点から、天候やワイパー操作については収録時の状況をそのまま採用するものとし、走行パターン関しても、東京都内の一般道路を走行すると言う条件で、もっとも自然な形になるようにした。ただし、それ以外のコントロール可能な要素については、窓閉、オーディオオフ、エアコンオフもしくは最弱と定めた

(2)について、図 1 で示すようにマイクは助手席の前、ダッシュボード上に直線状に 7 つ配置し、それらの間隔は右から順に 10cm, 5cm, 5cm, 5cm, 5cm, 10cm となっている。マイク番号も右から順に 1～7 となっている。平行して発話者に接話マイクを付け、これをマイク番号 8 とする。

音声データは、都心部を走行して収録し、男女 5 名ずつ計 10 名 1967 発話である。POI (Point Of Interest) 数は 152 個あるため、Julian において 152 個の POI 名を単語リストとして設定する。

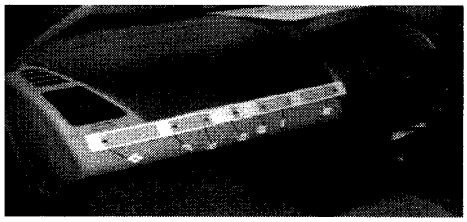

図 1. マイクの配置

(3) については、アプリケーションからの要請として、実際に使用されることが多い目的地入力、中でも重要度の高い POI(Point Of Interest)の入力を対象タスクとして採用した。

(4)は極めて重要な制限である。音声収録の方法としては、被験者に POI リストを与え、上から順に読み上げさせるというのが最も簡単な方法であるが、このようにして得られる読み上げ音声の性質は、しばしば実際のシステム入力と大きく異なっていることがある。そこで、実際のシステム入力に近いデータを得るために、被験者の自発性を引き出す収録方法が求められる。

3. 実験方法

Linux 上で音声データを Julian にかけ、認識率を出した。認識率は正解リストと照らし合わせ誤認識数をカウントして算出した。

①発話者に近いマイク 4 のデータで男女 5 名ずつ計 10 名、50 単語、計 500 発話の認識率を算出する。
②スペクトルサブトラクション（定常雑音成分の除去（以後 SS））を行い認識率を算出する。
③アレイマイクロフォンによる音声の加重合成を行い認識率を算出する。今回はフリーソフト Sound Engine を使用し、音声波形をそのまま加算した。発話者に近いマイク 4 のデータを中心に。
(1)マイク 3＋マイク 4
(2)マイク 4＋マイク 5
(3)マイク 3＋マイク 4＋マイク 5
を加算し、男女 5 名ずつ計 10 名、30 単語、計 300 発話の認識率を算出する。
④SN 比の算出(平均パワーの比)をする。

SN 比の示す値が大きいほど雑音が少なく、低ければ雑音の影響が多い。S/N 比は下記の式で求めた。

$$S/N\ 比=20\log10(Sp/Np)\ [dB]$$

4. 実験結果および考察

①の実験をしたところ、図 2 で示すように発話者毎に認識率が変化した。接話マイクの平均が 98.6%となり、マイク 4 の平均が 89.8%という結果になった。この結果から、マイク 4 の認識率をいかに接話マイクに近づけるかという課題が生まれた。そのために②③の実験を行った。

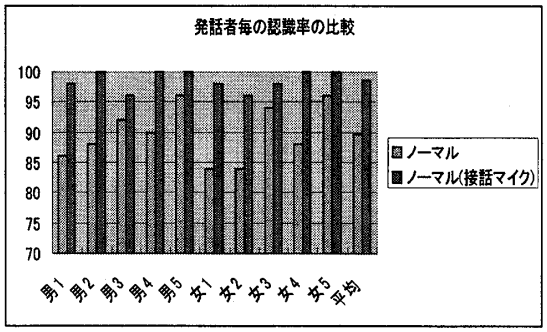

図 2.　発話者毎の認識率の比較

②の実験では、初めに音声波形ファイルからデータを読み込み、無音部を表示させ、発話者毎の無音部区間を測定した。そのデータを Linux 上で音声データを Julian にかけ、認識率を算出した。その結果、図 3 で示すように 10 名 50 単語平均で、92.6%と認識率の向上が見られ、接話マイクの平均に近づく結果となった。また、一律無音部 100ms の平均認識率 86.4%と比べても認識率が向上した。やはり、無音部設定が長すぎると音声と被ってしまい、短すぎると雑音推定が困難になるため、適応的に無音部を設定すると認識率が向上する結果となった。

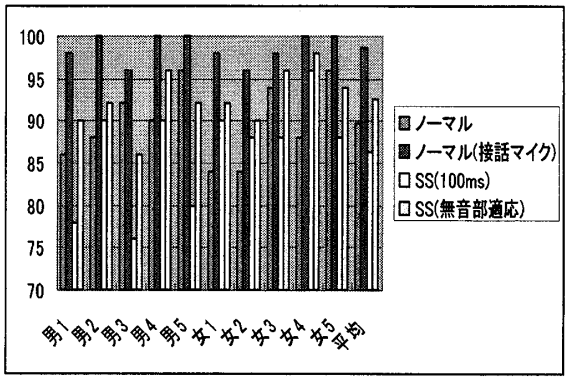

図 3. SS(話者毎の認識率の比較)

③の実験では、Sound Engine で加算させた音声を使用し認識率を算出した。その結果、図 4 で示すように男 2 以外、マイクを多く加算するにしたがって、少しではあるが接話マイクに近づく結果となった。しかし、音声波形を加算するので雑音まで加算され、大きく認識率が向上しなかった。

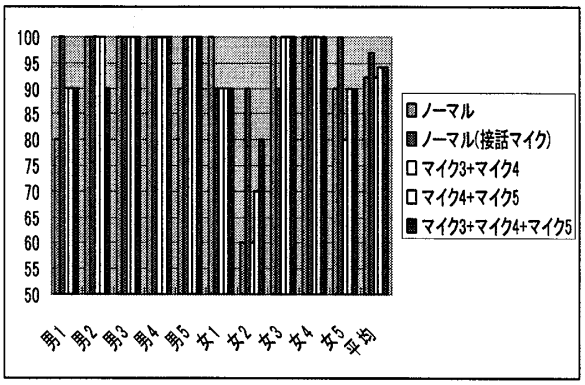

図 4. アレイマイクロフォン(話者毎の認識率の比較)

④の実験では、今回は何の加工も施していないノーマルの音声データを Power to Power 算出法を用いて算出した。表 1. で示すように、発話者毎において認識率が高い発話者の音声は SN 比も高く、認識率が低い発話者の音声は SN 比も低くなるという結果になった。

	S/N 比[dB]	認識率[%]
男 1	8.98	86
男 2	9.87	88
男 3	10.83	92
男 4	11	90
男 5	15.24	96
女 1	9.02	84
女 2	8.27	84
女 3	13.71	94
女 4	10.2	88
女 5	12.52	96

表 1.　発話者毎の SN 比の比較

5. まとめ

今回、接話マイクの認識率にどれだけ近づけるかというのを目標に実験を行った結果、SS では無音部長を適応的に変更することで、認識率が向上した。今後の展開として、アレイマイクロフォンの加算方法を精密に検討し、さらに SS 処理との併合を試みて認識率向上を目指していきたい。

謝辞

音声データは(株)日立製作所大淵氏から入手した。感謝します。

参考文献

[1]Julius　http://julius.sourceforge.jp/
[2]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」C-3～C-52　(平成 18 年 3 月)
[3]齋藤康夫他：「車載音声の解析と評価」平成 21 年東北地区若手研究者発表会 pp. 131-132(平成 21 年 2 月)