



## バイク走行環境下での音声認識／ 音声対話システムの性能向上に関する検討

Investigation into Improved Capabilities of Voice recognition/Voice response Systems in a Motorcycle Operating Environment

赤坂 貴志

### Abstract

As an interface for the operation of information tools while a bike is being ridden we are aiming to implement a voice recognition that is robust even in the noisy environment when running at about 100km/h. We planned to create acoustic models, which matches the high-speed operating environment, and the results of analyzing the sound recorded in this environment showed the noise was predominantly wind noise, which we believe can be simulated in a wind tunnel. At present we have built acoustic models constructed from sound data recorded in a wind tunnel and are performing voice recognition on it, obtaining better recognition results than off-the-shelf software in both the wind tunnel and when actually running at high-speed. We are proposing a music retrieval system as a motorbike information tool, based on this voice recognition equipment. We are investigating the demand for this system and its practical usability, including measuring the ratio of successfully completing tasks. The task still remains to further improve the performance of voice recognition and the need to create a music selection system that is specialized for bike use.

### 要旨

バイク走行中の情報ツールを操作するインターフェースとして、100km/h程度の高速で走行したときの雑音環境下でも頑健な音声認識を実現することを目指している。方針は、高速走行環境にマッチした音響モデルを構築して対応することであるが、同環境で収録した音声を分析した結果、この場合の雑音は風切り音が支配的であって、風洞でシミュレートできることが推察できた。実際に、風洞環境で収録した音声データで構築した音響モデルを使って認識を行ったところ、高速実走行環境、風洞環境とも、市販認識ソフトより高い認識率を得ることができている。また、この音声認識装置を基に、バイクの情報ツールとして、楽曲検索システムを提案した。“タスク達成率”という尺度も含めて、このシステムのニーズ、実用性を検討している。一方で、音声認識のさらなる性能アップと、バイク用に特化した楽曲検索システムの作り込みの必要性なども課題として残った。

## 1 はじめに

現在、カーナビに代表されるように自動車(四輪)の情報化がめざましく進歩しているのに比べ、バイク(二輪)の情報化は遅れている状況である。バイクの情報化の観点で社内アンケートを実施したところ、全く必要がないわけではなく、むしろ、バイク用のナビゲーションシステム、走行中の音楽鑑賞、コミュニケーションツールなどに興味、あるいは、ニーズがある結果が得られている。これら情報ツールを走行中に操作する適当なインターフェースがないこともバイクの情報化が遅れている理由の一つであると思われるが、それには、自動車以上にアイズフリー、ハンズフリーのインターフェースが必要である。その手段として音声認識を考えることは自然であるが、一方、バイクではマイクが外気にさらされることが必然で、したがって、走行時の雑音環境も自動車以上に厳しく、既存の音声認識ソフトでは対応できない。そこで我々は、バイク走行時、特に100km/h程度の高速で走行したときの雑音環境下をターゲットとして、その環境でも頑健な音声認識装置を実現することを目指した。また、実際のアプリケーションとして、前述のアンケートでもニーズがあった、走行中でも音楽鑑賞ができるようなソフト・楽曲検索システムを提案し、作成した音声認識装置を実装して、アプリケーションも含めた実用性を検討した。以下に報告する。

## 2 単語認識

本稿で扱う音声認識は、コマンドベースの単語認識である。単語認識のシステムは、おおよそ図1のようなブロック図で構成される(場合によっては各処理が前後したり、包含されたりすることもある)。

雑音環境下における音声認識率の向上というテーマに対するアプローチとして、図1の中で、雑音抑圧の方法、雑音に頑健な特徴分析(特徴量)の

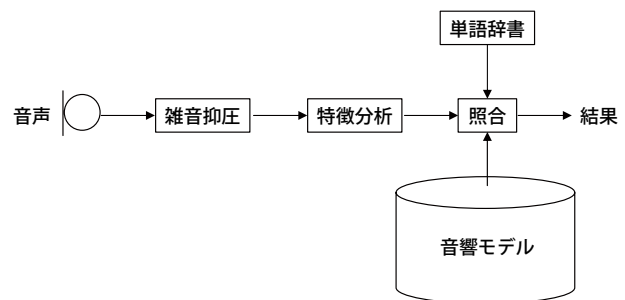


図1 単語認識ブロック図

検討、あるいは、照合時の雑音・話者への適応処理の方法などがよく研究されているが、本稿では雑音環境にマッチした、つまり、100km/hのバイク走行時の雑音環境にマッチした音響モデルを構築することで、全体の音声認識率の向上を目指した。ここで、音響モデルとは、入力された音声(の特徴量)がどんな音素(音素≡発音記号)であるかをまとめてあるような、いわば、音声の特徴をモデル化したデータベースのことで、大量の発話データを学習データとして構築される。

### 3 風洞による音声収録

音響モデルを構築するには、一般に、想定する環境において、一人当たり数十分程度のある決められた発話が、最低数十人規模で必要となる。すなわち、今回目指すところではバイクで100km/hの走行をしながら、上記の条件の音声データを収録しなければならず、現実的には容易ではない。ところが、ヘルメット内にマイクを装着(図2)し、実際に100km/hで走行したときに発話、収録された音声を分析すると、雑音としては風切り音が支配的で、音声認識にとって、他の雑音はほとんど影響していないように推測された。そこで、本稿では、風洞で30m/s(108km/h)の送風を行い、擬似的に100km/h走行環境を作り出して、この下で音声収録を行った。今回の被験者は男性話者105人、発話内容は、学習用データとしてATR音素バランス文<sup>1)</sup>10セットのうちいずれか1セット(2回ずつ)と、認識テスト用の15種類の単語(3回ずつ)である。



図2 ヘルメット内に装着されたマイク

## 4 認識実験

### 4.1 音響モデルの構築

被験者105人中70人分の学習用データを用いて、音響モデルを構築する。構築はHTK<sup>2)</sup>のコマンドに準じて学習した。

### 4.2 風洞データの認識

4.1で音響モデルの構築に使わなかった35人分の単語発話データを、上記音響モデルを使って認識した。単語は、

電話 はい いいえ キャンセル 会社 自宅 メモリーダイヤル こめ  
シャープ 詳細 現在地 到着予定時間 <個人名1> <個人名2> <個人名3>

の15種類、辞書の単語数は1000単語である。同じ単語発話データを、車載用音声認識として最も広く普及しているA社の音声認識ソフトでも認識させたので、**図3**にあわせて示す。

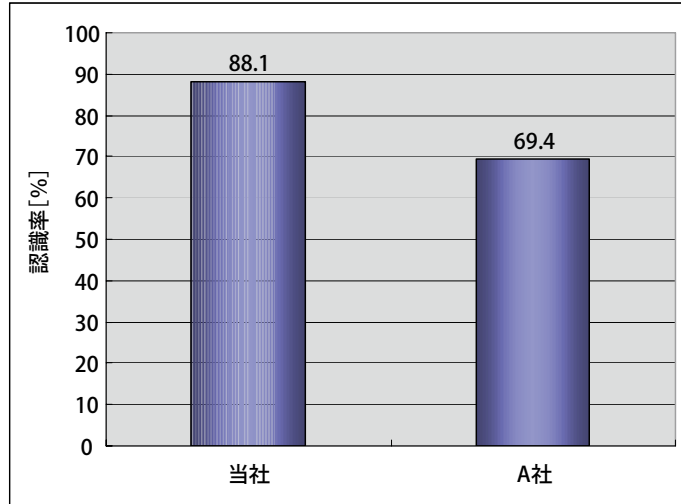


図3 風洞における音声認識率

#### 4.3 実走行データの認識

同一の音響モデルを使って、実際に100km/hで走行したときに発話した音声データを認識させた。話者(被験者)は一人で、単語は**4.2**の単語のうち、

電話 キャンセル 自宅

の3単語のみであるが、辞書は同じもの(1000単語)を使った。走行時はほぼ無風、極力100km/hになったのを確認して発話してもらった。**図4**に認識結果を示す。

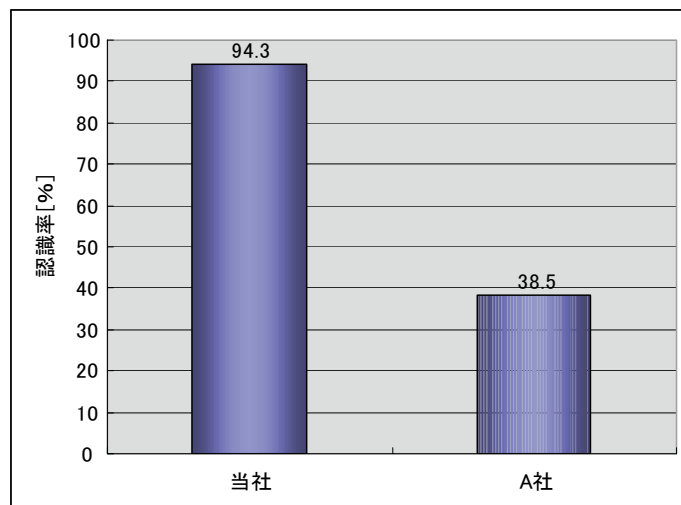


図4 100km/h 実走行における音声認識率

#### 4.4 結果

図3は市販の音声認識ソフトより高い認識率を上げていることから、使用環境で収録した音声を使って音響モデルを構築することにより、その認識率を向上できることを示している。また、図4をみると、対象話者・対象単語は少ないものの、実際の走行時の認識でも、市販の音声認識ソフトよりもはるかに高い認識性能を示しており、その認識率も図3に近く、風洞で擬似的な走行状態を作ることが妥当であることが分かる。さらに、これらは、本稿で目指している環境下において、市販されている音声認識ソフトでは、根本的にカバーしきれないことを明示している。

## 5 バイク用音声対話アプリケーション

### 5.1 楽曲検索システム

バイクの情報化に関するアンケートに音楽鑑賞のニーズがあったため、バイク用音声対話の実アプリケーションとして楽曲検索システムを作成した。これは、あらかじめ作っておいた楽曲データベースから聴きたい曲を、曲名、あるいは、アーティスト名を発声、認識することにより選曲・再生できるものである（選曲例：図5）。楽曲検索システムは、音声認識部にJulius<sup>3)</sup>を使った名古屋大学が開発したシステムであるが、発声トリガー用の押しボタンをつけて、発声タイミングを明示することなどで、バイク用にアレンジした。

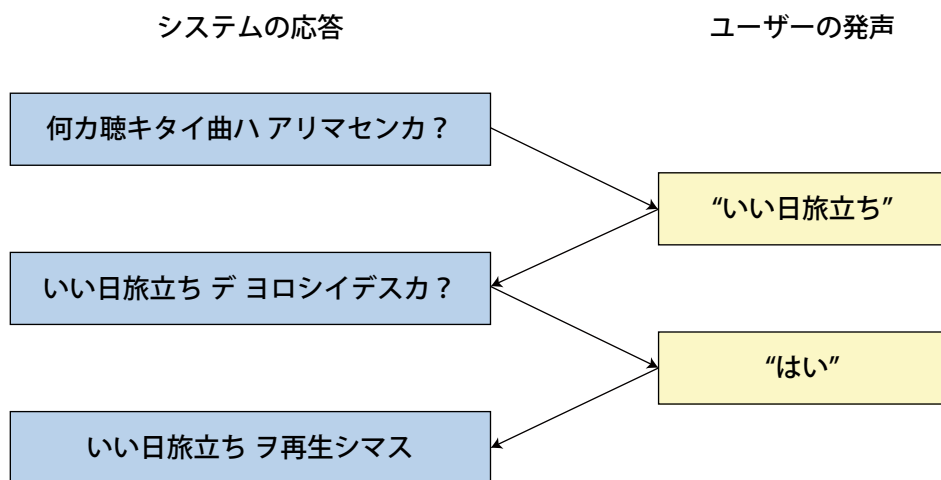


図5 音声対話による選曲の例

### 5.2 風洞におけるテスト

風洞にて、前項の楽曲検索システムのテストを行った。音声認識部(Julius)において、Julius付属のオリジナルの音響モデルを4.1で作成した100km/h走行用音響モデルに置換した。データベースは、曲数・約450曲、アーティスト数・約300人(グループ)で、この中から聴きたい曲を音声で選曲する。被験者が音声コマンドで選曲する行為を“タスク”とし、基本は、“聴きたい曲が聴けた”ことを“タスクの達成”、聞けなかったときを“タスクの未達成”とするが、1つのタスクの中で、同じ音声コマンド(曲名、アーティスト名含む)を3回以上誤認識した場合は、最終的に聴きたい曲が聴けても“タスクの未達成”とする。＜聴きたい曲が再生された数＞/＜聴きたかった曲(タスク)数＞×100[%]をタスク達成率として評価し、目標を80%以上とした。

図6に結果を示す。

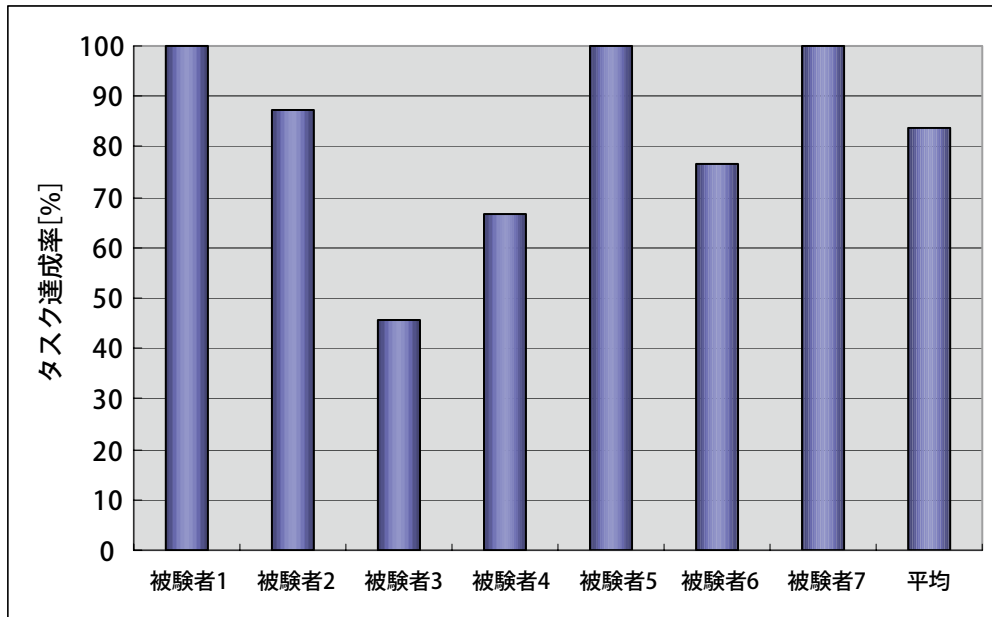


図6 風洞における楽曲検索システム：タスク達成率

### 5.3 実走行テスト

楽曲検索システムの実走行での操作性も含めた検証をするため、市街地を走りながら同システムをテストした。Grand Majesty400を使い、音声認識開始のトリガーを与える押しボタンを、図7のように、左ブレーキ上に設置した。音響モデルは、Juliusオリジナルのものを使用した。結果は被験者10人に対し、タスク達成率の平均90.7%、(被験者単位で)最低75.0%であった。



図7 音声認識開始トリガー用押しボタン<sup>4)</sup>

### 5.4 結果

風洞、実走行環境、どちらのテストでもタスク達成率の低い被験者は存在するが、平均としては目標の80%を達成することができ、ある程度の成果はあったように思われる。今後は、このようなタスク達成率の低い被験者の率の底上げが課題となる。風洞よりも、市街地走行の方が、全体的にタスク達成率が良いが、これは市街地走行では風切り音の影響が少なく、雑音的にクリーン環境に近いからだと思われる、これは、4.1で作成した音響モデルが、100km/h走行環境下では他のどんな音響モデルよりも高い性能を示すものの、「クリーン環境でのJuliusオリジナル音響モデルの性能」までは達していないことを表している。また、利用した名古屋大学製作の楽曲検索システムは、もともと画面を見ながら操作するアプリケーションで、その画面をみることで発話のタイミングが容易に分かるようになっている。今回は画面を見ずに操作できるように改良したつもりであったが、それでも、発話のタイミングがとれずに誤認識したケースも多かった。タスク達成率の向上には、音声対話システムも含め、バイク用にもっと特化した作り込みが必要である。

なお、5.3において被験者に、実験後、アンケート(意見、感想など)を実施したので、主なコメントを末尾に付す。

## 6 おわりに

バイクの情報化という観点から、その情報ツールをコントロールするインターフェースとして音声認識に注目した。むき出しマイクに高速の風が当たる中での音声認識ということもあって、最悪のシチュエーションであったが、音声認識にとって、バイクの高速走行環境の雑音は風切り音が支配的で、それは風洞でシミュレートできることが分かった。さらに、そのシミュレートされた環境で収録した音声データから構築した、いわば環境にマッチした音響モデルを根拠に音声認識装置を組み上げることで、その環境において、市販されている認識ソフトの性能をはるかに上回ることが明らかになった。また、実アプリケーションとして楽曲検索システムを提案し、実際の走行における操作性や、100km/h走行相当の環境でも実用レベルに近いことを示した。今後は、音声認識技術としては、100km/h走行(相当)時の音声認識率のさらなる向上、100km/h以上の風雑音環境での対応、女性話者の検討などと、これらと並行して、バイクに特化した音声対話システムの作り込み、専用機化・小型化の検討を行う必要がある。これらの技術を基に、ナビゲーションやコミュニケーションツールなど、さらなる展開を望むこともできる。

## 謝辞

本内容は、名古屋大学と共同で行ったものであり、同大学・武田一哉教授と武田研の学生に甚大な協力を受けたものである。ここに感謝の意を表する。

## 付録

市街地走行における楽曲検索システムに対するアンケート結果(意見、感想など)

- ・予想以上に聞き取ってくれた
- ・音声認識に感動
- ・市街地程度なら使えそう 慣れれば大丈夫
- ・音楽を聴きながら走るのも楽しいと思った
- ・実験は純粋に楽しかった 早く実現されれば良いと思った
- ・暑い!(音声認識の関係上、シールドを閉めなければならぬため)
- ・慣れないと操作は面倒
- ・検索中、信号が変わるとあせる
- ・交差点などうるさい場所では認識率が良くなかったような気がする
- ・機能を絞っても良いのではないか

## ■参考文献・引用文献

- 1) <http://www.jipdec.jp/chosa/public/report/onseidb/>
- 2) <http://htk.eng.cam.ac.uk/>
- 3) <http://julius.sourceforge.jp/>
- 4) 中田諭志, 原直, 赤坂貴志, 武田一哉 “バイク走行環境下での音声対話システムの性能向上に関する検討”, 日本音響学会 2007年春季研究発表会 ポスターセッション

## ■著者



赤坂 貴志  
Takashi Akasaka

コーポレートR&D統括部  
システム技術研究部