

車載用音声認識技術

Voice Recognition Technology for Car-Mounted Devices

岩見田 均 *Hitoshi Iwamida*
大和 俊孝 *Toshitaka Yamato*
崎山 和広 *Kazuhiro Sakiyama*
北尾 英樹 *Hideki Kitao*
高島 淳行 *Atsuyuki Takashima*

要 旨

車載機器の高機能化・複雑化が進む中、ドライバ - に安全で使い易いヒュ - マンインタ - フェ - スを提供する手段として、音声認識が注目されている。

今回、我々は車の様々な騒音下で良好な認識性能を確保するため、音声入力部と音声認識処理部の両面から検討を加えた。

本稿では、マルチマイク方式による入力信号の S N 改善や安定した音声の特徴抽出による認識率の向上、さらに騒音による誤動作低減を可能にした音声認識の処理方法について紹介する。

Abstract

With car-mounted devices increasing in terms of their function and sophistication, voice recognition is attracting attention as a means to provide drivers a safe, easy-to-use interface with the devices in their vehicles.

In a bid to achieve sufficient voice recognition performance in various noise conditions, we have considered the issue in two separate research areas: voice-input and voice recognition processing.

This paper describes the way a multi-microphone system improves the SN quality of input signals, the way the rate of recognition is raised by identifying steady voice characteristics, and a voice recognition method that lessens the likelihood of noise-induced malfunction.

1. はじめに

近年、ナビゲーションなど自動車の搭載機器の高機能化、複雑化や携帯電話の普及により、自動車運転中のこれらの機器の操作や画面表示の注視による安全性の低下が懸念されるようになってきた。

また、運転中に複雑な操作を行えないように装置側で走行中に使える機能の限定など行っており、必要な時に必要な情報を得られないなど機器が持つ本来の機能を十分に活用できていないのが実状である。

このような車載機器の問題を解決する手段として、音声認識を用いたヒューマンインターフェースが注目されている。

音声認識により、視線移動や手の動作を軽減させた操作が可能になると共に、これまで走行中に禁止されていた煩雑な操作を簡単な操作で実現できると考える。

我々は、この度、車の中の様々な騒音環境下でも高い認識性能を確保できる耐騒音音声認識技術の開発を行ったのでここに紹介する。

2. 開発のねらい

2.1 音声認識の技術動向

音声認識の車載機器への適用は、ナビゲーションの基本操作やハンズフリー自動車電話のメモリ発信操作を小語彙の特定話者音声認識を用いて行うことにはじまる。

現在では、ナビゲーションの目的地を地名や電話番号で設定するために、同時認識語彙数が数千単語の不特定話者音声認識に移行しているが、車の中という様々な騒音環境下で十分な音声認識性能を確保することが課題となっている。(図-1)

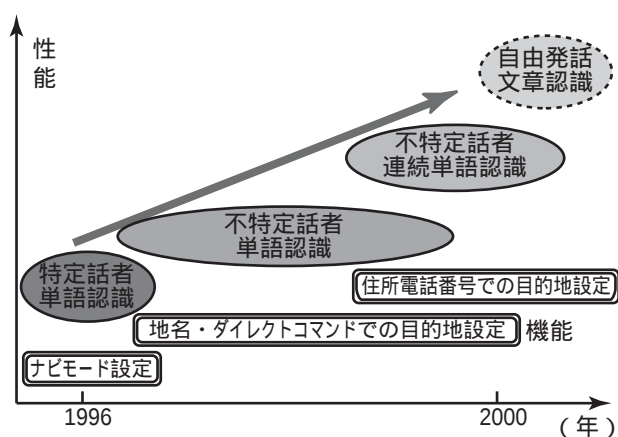


図-1 車載用音声認識技術の動向

Fig.1 Trend of voice recognition technology for vehicle

2.2 当社の取り組み

車の中では、ロードノイズやエンジン音などの走行騒音やウinker・ワイパーなど補機類からの騒音など様々な騒音が発生する。(図-2) 高速走行などで生じる大きな走行騒音は、音声信号(S値)と騒音信号(N値)のSN比を悪化させ、音声認識の誤り率を高めてしまう。また、ウinker音やワイパー動作音など非連続的な音は、音声認識が音声信号と誤って判断し、誤認識や不都合な動作へと繋がる。

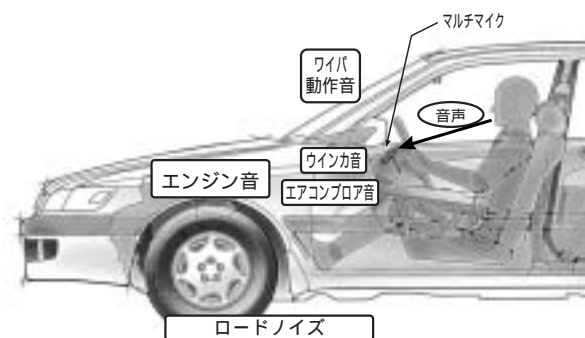


図-2 車室内の騒音

Fig.2 Noise in vehicle cabin

そこで我々は、認識率の向上を実現するため、マイクロホンの指向特性を鋭くすることにより、音声認識の入力部分で音声信号への騒音の影響を低減できる方式の開発を行った。

また、車の中では、様々な騒音を聞きながら発声するため、通常に比べて張り上げ気味の発声を行ってしまう。このため、オフィスなど比較的静かな環境下での発声とは異なってくる。そこで、音声認識の中心となる基本処理部での認識率を向上させるため、車の中で収録した大量の音声データを元に音声の特徴抽出能力を高めたアルゴリズムの開発を行った。

それと共に、ウinker音やワイパー動作音など補機類の発生音による音声認識の誤動作を低減するため、補機類の発生音と音声の特徴の分析に基づいたアルゴリズムの開発を行った。

2.3 音声認識の概要

音声認識の初段には、音声入力部が構成されており、マイクロホンで収録された話者の音声信号を音声認識処理部へ出力する。

音声認識処理部は、大きく音響処理部と単語照合部から構成されており、音響処理部は、入力信号の音声区間

を検出する分析部と音声区間の信号がどのような音素成分を含んでいるかを分析する音素照合部から出来ている。音素照合部には、様々な発声パターンを統計的に処理して導出された音素辞書を備えている。(図-3)

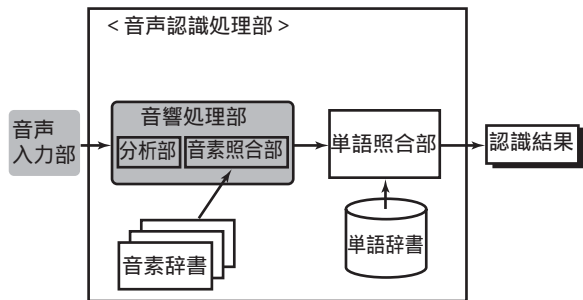


図-3 音声認識処理の概要
Fig.3 Outline of voice recognition processing

次の単語照合部は、音響処理部より出力される音素情報を元に単語辞書に登録された単語との比較・照合を行うことで最も確からしい単語の抽出を行う。

単語辞書は、テキストで記述可能のため、アプリケーションに合わせて書き換えが簡単に行えるようになっている。

3. 耐騒音音声認識技術の開発

本章では、認識率改善のための音声入力部での騒音低減と音響処理部で安定した音声の特徴抽出及び騒音による誤動作低減のための取り組みについて述べる。

3.1 音声入力部での騒音低減

3.1.1 ねらい

車載用音声認識の認識性能を高めるには、走行中に発生するロードノイズなどの騒音環境下でのS/N比の確保が必須である。

そこで、我々は、音声入力部に鋭い指向特性を持たせることにより、音声到来方向の入力感度を高め、それ以外の騒音が到来する方向の入力感度を下げることによってS/Nの改善を図った。

また、車載用の音声認識以外のシステムへも利用可能とするため、音声認識の入力部分(マイクユニット)で指向制御が完結することと、広い周波数帯域でフラットな周波数特性を得ることを前提とし、コンパクトで低コストのマイクユニットの実現の為にアナログ回路方式で開発を進めた。

3.1.2 マルチマイク方式の原理

図-4にマルチマイク方式の構成ブロック図を、図-5に指向特性生成の模擬図を示す。

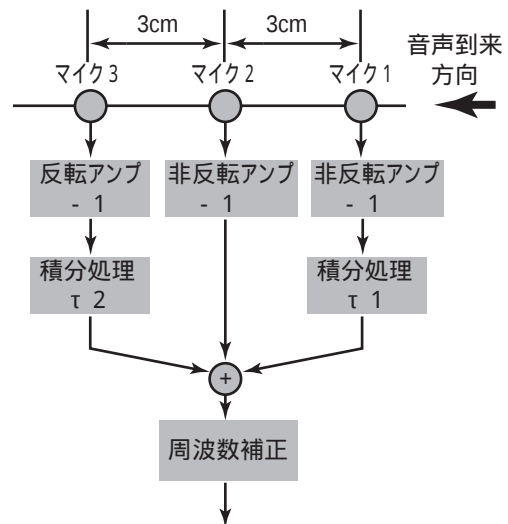


図-4 マルチマイクの構成
Fig.4 Constitution of multi-microphone system

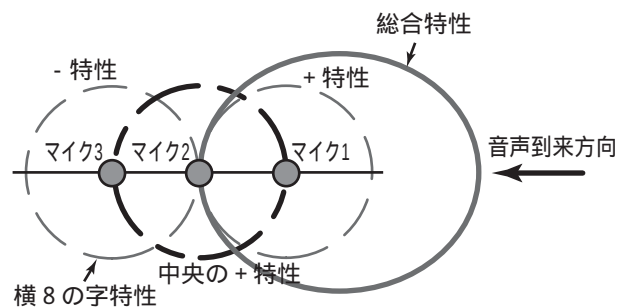


図-5 マルチマイク指向特性の模擬図
Fig.5 Simulation drawing of multi-microphone directivity pattern

以下に、所望の指向特性を実現する具体的な処理内容を説明する。

- 1) 3つの無指向性マイクロホンを一定間隔に直線配置
 - 2) 両端のマイクロホン（マイク1、マイク3）の位相差を逆位相とし、積分処理を行ない加算することで、中央のマイクロホン位置を中心に「横8の字」の特性を生成
 - 3) 中央のマイクロホン（マイク2）の正位相を加算することで、マイク1の方向に指向特性を生成
- 尚、マイク1とマイク3の出力部にある積分処理は、積分器の持つ低域増幅特性を利用して、「横8の字」特性が得られにくい低周波数域の特性改善を行うものである。また、周波数補正部は、4kHz以上の高周波数域での指向

特性の劣化をローパスフィルタの共振特性を利用して補償するものである。

これら処理によって、マイク3側の感度を低下させ、マイク1側をより強調して、音声到来方向に指向性を持つ特性を実現した。

3.1.3 指向特性による効果

図-6に今回製作したマルチマイクによって得られた指向特性を示す。

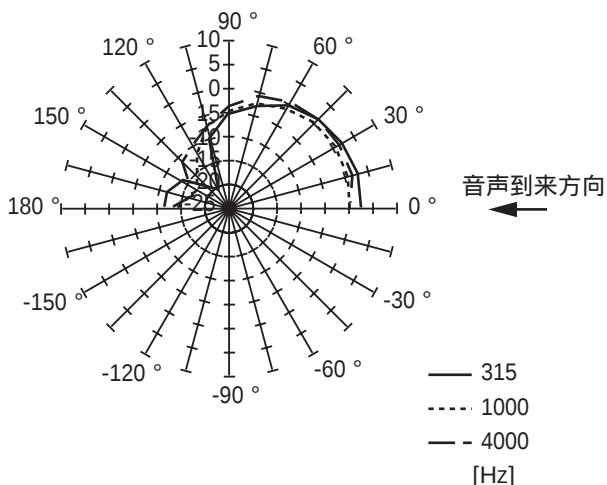


図-6 マルチマイク指向特性
Fig.6 Directivity pattern of multi-microphone system

300Hz付近から5kHz付近までの広い周波数帯域で、音声の到来方向(0°方向)に対し、それ以外の方向(135°～180°方向)で15dB以上の感度差のある指向特性を実現した。

図-7に上記の指向特性を持つマルチマイクによる実車での騒音低減効果例を示す。データは、高速100km/h走行時に「サッポロ」と発声したときの従来の1マイクとマルチマイクの音声波形である。車両走行時にさまざまな方向から到来する騒音のレベルが、マルチマイクの指向特性によって低減されていることが音声区間の前後の騒音波形より確認できる。

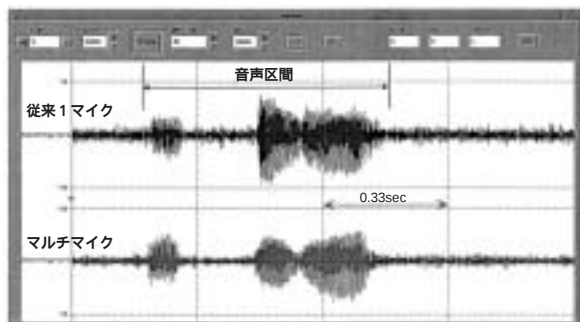


図-7 高速走行時の波形比較
Fig.7 Comparison of wave form during high speed running

なお、マルチマイクのマイク取付位置の最適化検討を行った結果、従来のマイク取付位置であるAピラー中央やサンバイザに比較して、ステアリングコラムカバー上部への取付が指向特性のより高い効果を得るために最も適していることが判った。

今回開発したマルチマイクを写真で紹介し、その特性を表-1に示す。



表-1 マルチマイク特性

方式	3マイクアレーを用いたアナログ処理
構成	指向性制御回路を内蔵した小型ユニットタイプ
サイズ	95mm × 30mm × 15mm
指向特性 (1kHz時)	135°方向 / -18dB 180°方向 / -15dB * 0°方向(音声到来方向)を基準
周波数特性	300Hz ~ 5kHz (0 ± 3dB / 0°方向)

3.2 安定した音声の特徴抽出

車載用音声認識では、車の中で発生する様々な騒音下で安定した音声の特徴抽出を行う必要がある。そこで、今回は、アイドリング状態と高速走行状態で収録した62名の音声データを用いて、以下に述べる音声認識処理部の開発を行った。

3.2.1 ケプストラムの採用

従来、音声の特徴を抽出する手段としてパワースペクトラムを用いていた。パワースペクトラムは、音声波形をフーリエ変換することで周波数領域に展開し、パワー計算を施したものである。

今回採用したケプストラムは、対数処理したパワースペクトラムを、逆フーリエ変換を施したものである。(図-8)

パワースペクトラムとケプストラムの大きな違いは、図-9の様にその微細構造にあり、パワースペクトラムでは音声入力の微妙な差が微細構造に反映されてしまうため、音素辞書とのパターン照合に悪影響を及ぼす。一方、

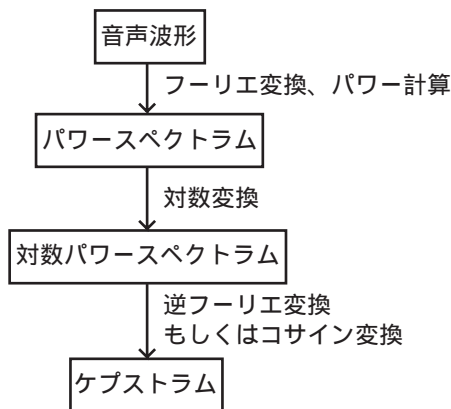


図-8 音声波形と各特徴抽出処理の関係

Fig.8 Relation between sound wave form and characteristic extraction processing

今回採用したケプストラムは、スペクトルの包絡（概形）のみを取り出す事が可能であり、発声の違いや騒音などで音声の特徴が微妙に崩れた場合にも比較的安定した音素照合を可能とした。

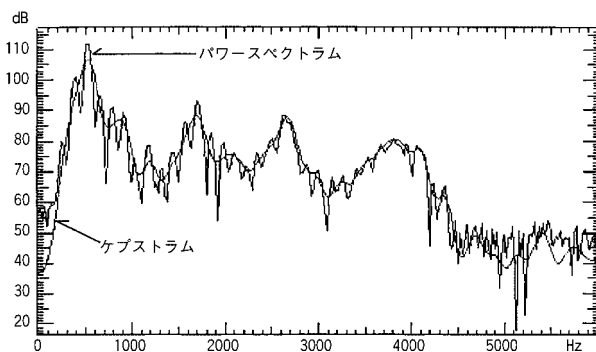


図-9 パワースペクトルとケプストラムによるスペクトル包絡の比較

Fig.9 Comparison between power spectrum and supectrum envelope of cepstrum

3.2.2 分析窓長の拡大

音声の特徴抽出は、一定周期で一定の幅を持つフレーム単位で行われ、この一定の幅をもった区間を分析窓という。一般的に分析窓長を長くすると周波数領域での分解能は高くなるが、時間的な変化は捉えにくくなる。

今回、実験的に調整を行った結果、音声の持つ周波数的特徴抽出を重視した方が、認識性能に良い結果が見られたので、従来に比べて分析窓長を長く設定した。

3.2.3 ケプストラムの重み付け

音声の特徴をケプストラム領域で見ると、低域では雑音の影響を受けやすく、高域では音声認識に有効な成分が少ない傾向にある。

そこで、騒音に阻害され難く、音声認識に有効な音声の特徴を含んでいる中域を強調するケプストラム領域での重み付け処理を行った。

3.3 騒音による誤動作低減

車室内では、走行騒音のほかにウィンカやワイバの操作音や動作音が発生し、音声の区間検出部で誤動作を起こしやすい。そこで、6車種の機器操作・動作音および数字発声（イチ、ニ、ヨン、ゴ、ナナ）の継続時間に着目し、調査を行った。（図-10）

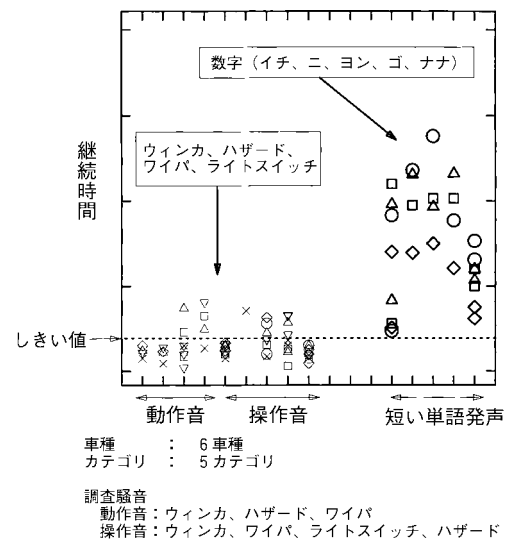


図-10 機器騒音と音声の継続時間

Fig.10 Duration of apparatus noise and voice

この結果、図-10に示すしきい値付近で車載機器の騒音と比較的短い音声を分離することが可能と判る。

一方、人間が発声する音声の中には、「コップ」の「ッ」のような閉鎖音や無声化による無音空間を持つものがあり、その前後では単音が孤立している。例えば、前の例では「コ」と「プ」が独立しているように見える。このような場面において、前後の単音は、単語として発声する単音よりも短くなる傾向があり、しきい値以下の信号を騒音と扱うのは危険である。

そこで、「短い単発の信号は騒音として扱う。但し、前後一定期間内に規定パワー以上の信号が入った場合は、音声と見なす。」という処理を実装した。前後の一定期間はウィンカやワイパーなどの音の間隔未満になるように設定され、音声との分離を実現している。

アプリケーションによっては、単音の単語が全くないことも考えられる。また、ユーザーに単音の発声を比較

的ゆっくり発声させることが可能であれば、騒音と扱う信号の幅を更に広げることで、より頑強なシステムが実現可能である。

4．まとめ

前章で報告した耐騒音性能を高めるための3つの方式を適用することで、次のような音声認識性能の改善と誤動作低減を実現した。

音声入力部で騒音低減が行えるマルチマイクと安定した音声の特徴抽出が行える音響処理部の開発の組合わせにより、不特定話者の音声認識として同時認識語彙数1000単語、高速100km/h走行の条件下で90%以上の認識率を得ている。また、騒音による誤動作を低減できる音響処理部の開発により、誤動作率を10%程度まで減少することができた。

以上の様に、今回開発した音声認識は、現在、市場に投入されている他社の製品と比較しても充分な性能レベルに達したと判断しており、近い将来に商品化を実施したい。

筆者紹介

岩見田 均(いわみだ ひとし)



1983年株式会社富士通研究所入社。以来音声認識の研究に従事。1988年～1991年株式会社ATR視聴覚機構研究所に出向。

5．おわりに

音声認識を車の中で使用する場合の一番の課題である耐騒音性能向上に関する当社の取り組みを紹介した。

今後も高機能化・複雑化する車載機器を車の中で使い易くする方策として音声認識は有望であるが、単純にスイッチの置き換え的な使い方では、本当の使い易さは得られない。

また、パーソナルコンピュータの分野では、読み上げた文章をワープロへ入力できるアプリケーションが登場するなど、連続単語認識の開発が活発になって来ている。

今後は、更に耐騒音性能の向上を図るとともに、連続単語認識の開発を進め、様々なしゃべり方に対応した音声入力の実現により、ユーザがストレスなく快適に車載機器を操作できる新しいヒューマンインターフェースの構築を目指したい。

大和 俊孝(やまと としたか)



1985年入社。以来音響技術の開発を経てデジタル信号処理技術の開発に従事。現在技術開発部在籍。

崎山 和広(さきやま かずひろ)



1991年入社。以来デジタル信号処理システムの開発に従事。現在技術開発部在籍。

北尾 英樹(きたお ひでき)



1992年入社。以来デジタル信号処理技術の開発を経て音声処理技術の開発に従事。現在技術開発部在籍。

高島 淳行(たかしま あつゆき)



1984年入社。以来音響システム、デジタル信号処理システムの開発に従事。現在技術開発部次長。