航海支援システムにおける音声認識に関する研究

松田 和生*・宮原 末治**

Study on Speech Recognition in Bridge Operation Support System

Kazuo MATSUDA and Sueharu MIYAHARA

Abstract

A new type of bridge operation support system has been developed in the project for developing a modernized coastal service vessel, which started on the background of reduction of Japanese sailors and shortage of young personnel resources in domestic marine transportation. In the present, this system is in practical use on various domestic coastal service vessels.

This system is aiming at ensuring navigational safety at one-person bridge operation, and has human machine interface with speech communication applied to prevent a watch keeper from mis-operations of the system and enable him to easily operate the system without stress.

It is essential and necessary to increase accuracy of speech recognition in order to reduce the burden of a watch keeper in operating a ship by means of speech communication function of bridge operation support system.

In this paper, study on improvement of speech recognition rate by using plural speech recognition software on the market is described.

Keywords : bridge operation support system, human machine interface, speech recognition

キーワード: 航海支援システム, ヒューマンマシンインタフェース, 音声認識

1. はじめに

内航海運では、船員の高齢化と厳しい労働環境による若年層の船員離れ等を背景に、平成 5 年度から全国内航タンカー海運組合、運輸省船舶技術研究所（現海上技術安全研究所）、学識経験者、造船所等が共同で内航船近代化研究を実施し、従来より少ない乗組員でも安全かつスムーズな運航が実現できる内航タンカーの検討を行った。

航海部門では、「船橋一名当直でも当直員の負担を増やすことなく、安全性を担保する」ことを目標に航海支援システムの開発を行った。開発されたシステムは、平成 9 年の初号機搭載から平成 9～10 年の安全性評価を経て、現在まで複数の内航船に搭載されて、有効に活用されている。

本システムは、自動操船機能、避航操船機能等を含めた高度な支援機能と音声を利用した人に優しいヒューマンマシンインタフェースを特徴としており、航海当直員からの操船指令、航行状況問い合わせ、

確認応答、電子海図表示画面切替等のコマンドを音声認識することで、操船及びシステムの主要な操作を音声で実行できるように設計されている。これにより、航海の種々の局面で航海当直員はシステムと会話を行うことによって、あたかも操舵員、レーダー等監視員と共に当直を取っているかのように、必要な情報収集、的確な操船を行うことができる。これにより、システムの使用に対する負担の大幅な軽減が可能となる。

反面、航海当直員の発する音声コマンドの誤認識が多くなると、逆に音声による指令、操作が負担となり、航海当直員の余裕の減少から安全性が損なわれる危険性もある。そのため、システムの音声認識精度を一定レベル以上に保持する必要がある。

現在搭載されている航海支援システムにおいては、音声認識装置に市販の専用機（音声認識用信号処理ボードとソフトを組込み）を使用し、90%以上の認識率を達成している。加えて、誤り訂正を考慮した

* 正会員 三菱重工業株式会社長崎造船所 (〒850-8610 長崎市飽の浦町 1-1)

** 非会員 長崎大学工学部 (〒852-8521 長崎市文教町 1-14)

音声認識率の向上対策、誤認識時のフェイルセーフ処理を導入して、システムとしての信頼性を高いレベルで維持できるように考慮している。

今後、認識系自体の性能向上により認識率アップは期待できるが、本研究では、現状入手可能な認識系を使用することを前提に、更なる音声認識率の向上を目的として、市販の音声認識系をシステムとして複数個適用し、音声認識精度を向上させる手法について検討する。⁽¹⁾

2. 音声認識を適用した航海支援システム

2.1 システム概要

船橋一名の当直体制では、航海当直員の業務負担及び精神的負担はかなり大きくなる。船橋での快適な作業環境を確保し、航行の安全性を担保するには、航海当直員の業務を広い範囲で支援し、操作し易く分かり易い、人に優しいヒューマンマシンインタフェースを備えた航海支援システムの装備が一つの解となる。

平成5年度からの内航タンカー近代化研究における航海支援システムの開発過程では、幅狭海域及び狭水道航行における船橋での船長、操舵員、レーダ等監視員の3人を基本とした従来の運航体制に代えて、船長以外の2人の役割を航海支援システムが分担する当直体制を想定・モデル化した。その場合に航海支援システムが操舵員及びレーダ等監視員の役割を肩代わりできるように仕様を検討した上で、目視による見張りと周囲状況の確認、高度な判断が求められる船長の役割は、航海当直員が受け持つことができるように考慮した。

このように、人間、航海支援システムの役割分担を明確にして、自船航行情報、他船情報、自船周囲環境情報等の操船に必要となる各種情報の収集、当直時に発生する種々作業を可能な限り支援する機能を設計した。加えて、航海当直員にとって分かり易い形式への収集情報の加工、航海局面に応じた必要最小限の情報の自動提供、避航判断支援、定時運航支援等の高度な支援機能も航海支援システム側で構築し、航海当直員が最重要の当直業務である目視による見張りに余裕を持って専心できるようにした。

また、航海当直業務の有効な支援と航海当直員の余裕創出のためには、人間と航海支援システムが円滑に意志疎通できることが大前提であり、この観点から操作が負担とならないヒューマンマシンインタフェースの構築が必要不可欠となる。

航海支援システムのヒューマンインタフェースに対する要件は次の通りである。

- ・見張りを妨げないアイフリー操作が行えること
- ・簡単で習得し易いこと
- ・迅速で信頼性が高いこと

- ・必要最小限の情報を提供できること

これらの要件を満足させるため、本航海支援システムでは、音声入出力（音声認識・合成）機能を新たに導入した。航海当直員が必要な情報の収集、操船を音声による対話で実行できるように、次の音声入力用の語（単語、単文）を準備している。

- ・操船指令： 約 10 語
- ・問い合わせ： 約 10 語
- ・電子海図画面操作： 5 語
- ・応答・その他指令： 約 10 語
- ・数値： 桁読み（針路値、船速値等）
- ・その他： 10 語（テレグラフ分画）

本航海支援システムの構成図を Fig. 1 に示す。⁽²⁾

Fig. 1 Configuration of bridge operation support system

2.2 音声認識の信頼性向上

本航海支援システムへの音声認識適用においては、船橋での航海業務で使用されるため、次の条件を考慮する必要がある。

- ・ワッチ体制に基づき、複数の乗組員に交代で使用される。
- ・当直交代時には引き継ぎ時間はほとんどなく、個人の音声登録、設定変更、認識系の切換等は困難である。
- ・オフィス等に比べて騒音レベルの高い場所で使用される。

上記条件により、本システムでは、不特定話者対応で、入力音声と周囲雑音の2入力差分方式の音声認識・合成装置（音声入出力装置）を採用しており、実船環境下での評価で約92%の認識率を得た。

更に、航行の安全性確保のため、航海当直員、航海支援システムを含めた操船システムとして、次の対策、処理を施し、全体での信頼性の維持、向上を図っている。

- ・入力語（指令語）は航海の現場で通常使用されている用語とし、話者による個人差の出にくい単語または単文に限定する。
- ・事前に入力語の認識テストを実施し、認識率の悪い単語または単文を航海の現場で使用されている認識率のよい同義の単語または単文に置き換える。
- ・航海当直員からの音声操船指令に対しては、システムは必ず認識した通りに復唱し、誤認識に対する即座の誤り訂正を可能とする。
- ・航海当直員からの音声操船指令に対する誤認識については、即座に音声または画面タッチ操作による中断を可能とする。
- ・音声操船指令において、コース値、船速値等が未入力の場合には、入力を促す音声メッセージを出力する。
- ・操船制御は、航海当直員の確認応答“オーケー”、“了解”等を認識した後に実行する。この他に画面でのタッチ操作による確認も可能とし、確認操作の確実性を高める。
- ・音声操船指令、報告・警報等すべての音声指令と音声出力メッセージに優先順位を付け、音声出力中でも緊急を要する音声指令、音声警報が割り込めるようにする。

上記の対策・処理により、情報の授受のみならず、操船制御も音声で安全に実施できると評価されている。(3)(4)(5)

2.3 音声認識における課題

本航海支援システムでは、操船制御、航行状況問い合わせ等の音声による主要操作を機能要件としている。このため、音声認識精度が悪くなると、再発声、誤り訂正等が頻発し、スムーズな操作ができなくなる。この結果、航海当直員にストレスが積み重なり、安全性を大きく損なう危険性が出てくる。

音声認識精度を低下させる要因としては、次の項目がある。

- ・周囲の雑音
騒音の大きい環境下では、実際の音声入力信号に騒音に基づく雑音信号が重畳される。音声認識装置端で雑音信号が重畳されて乱れた音声信号を処理するために、認識精度が悪化する。
- ・音声入力者の違い
人により発音、発声の特性、話し方が異なるため、一般に、使用者の違いにより音声認識率が異なることが多い。
- ・認識系の違い
認識系の有する音声特徴抽出・照合処理、音響モデル等の違いにより、認識精度に差が出る。
- ・言語辞書の違い
認識対象の単語、単文等は言語辞書に標準パター

ンとして登録する方式が一般的である。音のよく似た単語を多数登録すると認識率は低下する傾向にある。

開発したシステムでは、前述の通り専用の音声認識装置を使用しているため、その装置の特徴により、周囲の雑音、音声入力者の違いに対しては、高い認識性能を示している。また、入力語については、使用する単語、単文に対する事前評価を行って、誤認識が最小限となるように単語、単文を厳選して対応している。

課題としては、船舶の船橋で複数の航海当直員が入れ替わり使用する条件において、認識系の特性に起因する誤認識の排除、認識対象の単語・単文を増加させた場合の認識精度維持（拡張性の保証）がある。

3. 複数の認識系を併用した音声認識システム

本論文では、認識系自体の特性に起因する誤認識の排除、雑音混じりの入力音声に対する認識精度の向上、航海支援システムの機能拡張に伴う言語辞書の語彙増加における認識精度維持を目標に、複数の認識系を併用して認識率を向上させる手法について検討した。

3.1 システム構成

音声認識率を向上させるために構築した音声認識システムを Fig. 2 に示す。

Fig. 2 Configuration of speech recognition system

システムでは、一つの音声入力に対して複数の認識系（ここではA、B、Cの3種を使用）で同時に認識を行い、それぞれの系の認識結果として複数の認識結果（出力候補群）を得る。この出力候補群に対して、統合処理部において評価処理を行い、最終的な出力結果を決定・出力する。

3.2 認識系

本論文で使用した認識系は、開発した航海支援システムで適用している音声認識装置（認識系Aとす

る。)に加えて、汎用性、使用の容易さ、拡張性、コストを考慮して2種の市販ソフト(認識系B及びC)を適用した。

認識系A及びBは、複数の認識結果を確からしき付で出力し、認識系Cは、最も確からしい認識結果を1個のみ出力する。

それぞれの認識系では、認識対象語の設定として、認識させたい単語及び単文をテキストで認識系固有の言語辞書に登録する方式であり、容易に追加、修正、削除が可能である。

3.3 統合処理

それぞれの認識系が出力する複数の認識結果(候補)から1つの出力結果を評価、決定するにあたり、次の4種の処理を検討した。

(1) 多数決判定処理

それぞれの認識系の特性から生ずる認識結果の差を吸収することを目的として、各認識系の認識結果から多数決により出力結果を決定する。

本判定処理においては、認識系A、B、Cを使用して、次の手順で出力結果を決定する。

- ① それぞれの認識系が出力する認識結果(出力候補群)の第一位候補を比較し、多数決で出力結果を決定・出力する。
- ② 多数決で決まらない場合には、出力結果は棄却される。

(2) 相対距離値判定処理

多数決判定処理において出力結果が棄却となる場合に、第一位以外の認識結果(候補)の中に正解が含まれていることが多い。また、認識が曖昧な場合には、複数の認識結果が同等の確からしさで出力されるが、第一位候補ではない候補が正解であるケースも多い。

このような場合に、複数の認識結果(候補)から1個の出力結果を決定する指標として、認識系A及びBが認識結果とともに出力する確からしさに着目する。ある候補の確からしさの数値から、他の候補と比較した上での相対的な認識信頼度を判断することができる。

しかしながら、認識系から出力される確からしさは個々の系における音響モデルに基づいた固有の値であり、そのまま異なった認識系の比較、評価には使用できない。

そこで次の相対距離値 D_n を定義する。

$$D_n = \frac{|S_n - S_1|}{S_1}$$

ここで、 S_1 : 第一位候補の確からしさ
 S_n : 第 n 位候補の確からしさ

上式で定義した相対距離値は、「ある候補がその認識系の第一位候補に比べてどの程度信頼性があるか(ある候補が第一位候補とどの程度距離が離れているか)」の指標であり、この値が大きいと認識結果としての信頼性は低いと言える。

また、相対距離値を導入することによって、異なった認識系間でも、同じ候補を基準としてその相対距離値を比較することで、認識系の特性を越えた評価を行うことが可能となる。

認識系A及びBにおける各認識結果の相対距離値を使用して、次の手順に基づきシステムの出力結果を決定する。

- ①-1 認識系毎にすべての候補に対して相対距離値 D_n を算出する。
- ①-2 認識系A及びBの相対距離値付の候補を相互に比較し、両系共通の候補を抽出する。
- ①-3 共通候補が一つであれば、その候補を出力結果として決定・出力する。
- ②-1 共通候補が二つ以上であれば、各候補の相対距離値を平均し、相対距離値の小さい順に出力候補順位を付ける。
- ②-2 出力候補順位の第一位を出力結果として決定・出力する。
- ③ 認識系A及びBで共通候補が無い場合には、出力結果は棄却される。

(3) 候補数判定処理

個々の認識結果を評価すると、誤りとなった認識の場合には、小差の確からしさで複数の候補が出力される一方で、正解となった認識の場合には、正解と誤りの確からしさの差が大きく、候補数も少なくなる傾向にある。

そこで、認識系の出力する候補の数に着目し、少ない候補を出力した場合にはその認識結果は信頼度が高いと判定する。

認識系A及びBを使用して、それぞれの認識系が出力する認識結果(候補)の数によって、次の手順に基づきシステムの出力結果を決定する。

- ① 候補が一つの場合には、その候補を出力結果として決定・出力する。
- ② 認識系A及びBの出力する候補がともに一つでその候補が異なる場合には、出力結果は棄却される。
- ③ 認識系A及びBがともに二つ以上の候補を出力している場合には、候補数の少ない認識系の第一位の候補を出力結果として決定・出力する。
- ④ 最少の候補数が同じであれば、それぞれの候補の平均順位により決定する。
- ⑤ 平均順位も同じであれば、出力結果は棄却される。

(4) 複合判定処理

認識系 A 及び B を使用して、相対距離値判定と候補数判定の複合した次の手順でシステムの出力結果を決定する。

- ① 候補が一つの場合には、その候補を出力結果として決定・出力する。
- ② 認識系 A 及び B の出力する候補がともに一つでその候補が異なる場合には、出力結果は棄却される。
- ③-1 認識系毎にすべての候補に対して相対距離値 D_n を算出する。
- ③-2 認識系 A 及び B の相対距離値付の候補を相互に比較し、両系共通の候補を抽出する。
- ③-3 共通候補が一つであれば、その候補を出力結果として決定・出力する。
- ④-1 共通候補が二つ以上であれば、各候補の相対距離値を平均し、相対距離値の小さい順に出力候補順位を付ける。
- ④-2 出力候補順位の第一位を出力結果として決定・出力する。
- ⑤ 共通候補が無い場合には、最終候補の相対距離値が大きい認識系の第一位候補を出力結果として決定・出力する。

⑤の処理については、正解と誤りの確からしさの差が大きい程、認識信頼度が高いことに基づく。

4. 実験及び考察

4.1 実験データ

前述の航海支援システムで使用している入力語は、指令、問い合わせ等の約 35 語の他に、針路値、船速値等の桁読み数値及びテレグラフ分画（ナビフル等）がある。本研究の実験では、将来の機能拡張を考慮して約 60 語を追加して、指令、問い合わせ等で 94 語、数値の代表例 46 語を含めて計 140 語を適用した。実験に使用した語を Table. 1 に示す。

4.2 実験条件

上記 140 語を成人男子 3 名が発声した音声を用いて実験を行った。発声にあたっては、事前の発声指導、練習は行っていない。

Table. 1 A number of words and simple sentences for speech recognition experiment

Contents	Words & Sentences	Examples
Maneuvering orders	32	「コース」, 「船速」, 「機関停止」, 「ステディ」など
Inquiries	32	「針路は」, 「潮流は」, 「危険船は」, 「座礁危険は」など
Display operations	17	「ヘッドアップ」, 「拡大」, 「縮小」など
Responses	13	「OK」, 「待て」, 「了解」, 「はい」, 「いいえ」など
Values	46	「0(ぜろ)」, 「0(れい)」, 「100」, 「350」, 「+」, 「-」など

※ Total 140 words and simple sentences

実験では、発声した個々の音声に 749GT 型貨物船が 100% 主機負荷で航走中に船橋で収録した音をミックスして使用した。

認識系はいずれも事前学習、話者適応は実施しておらず、認識系 A については、2 入力差分方式を用いた雑音除去機能は外して実験を行った。

4.3 実験結果

実験結果を Table. 2 に示す。

4.4 考察

認識系単独での認識率 90~92% が認識系を複数使用して統合処理を行うことにより、94~97% まで改善することができた。特に相対距離値判定と候補数判定を複合した処理で約 97% の認識率となり、単独認識系による認識率より約 5% 向上し、複数の認識系による効果を確認できた。

複合処理における誤認識の結果を分析すると、一つの認識系が誤りである候補を 1 個のみ出力したケースが約 43%、一つの認識系では正解を認識したが、もう一つの認識系では候補に上がらなかったケースが約 36%、正解より他の候補の相対距離値平均値が小さいケースが約 21% であり、少なくとも一つの認識系には正解が候補として出力されていた。

Table. 2 Results of speech recognition experiment

	Speech recognition rate (%)						
Processing	Individual			Decision by majority	Relative distance values	No. of proposed outputs	Composite process
System	A	B	C				
Correct	92.1	90.0	92.1	96.0	94.3	95.0	96.7
Reject	0.0	0.0	0.0	2.1	2.6	1.0	0.0
Error	7.9	10.0	7.9	1.9	3.1	4.0	3.3

これより、確からしさ付の候補を出力する2つの認識系を適用して適切な処理を施すことにより、認識精度を改善できることがわかった。

また、認識系Cを含めた3つの認識系で多数決判定処理した場合にも約4%の認識率向上が図れた。この場合には、誤り率は約2%であり、複数の認識系による多数決だけでもかなりの認識率向上が得られることがわかる。

5. まとめ

航海支援システムにおける音声認識に関し、その精度向上手法を検討したが、市販の音声認識系を複数使用して、同時に認識された結果に対して処理を施すことにより、船橋の騒音環境を模擬した条件においても、単体で使用した場合に比べて認識率を平均で約4%向上させ、誤り率を1/2~1/5に低減することができた。これにより、認識系の複数使用は、音声認識率向上の一手法として有効であることが確認できた。

最新のコンピュータ技術を適用すると、1台のコンピュータで複数の音声認識プロセスを動作させることは容易であり、高いパフォーマンスも確保できる。これにより、大幅なコスト増も無く、開発済みの航海支援システムにも適用可能と考える。

今回は、1種類の雑音環境と3名の成人男子による実験を行ったが、今後は、被験者、雑音環境及び入力語を増やして実験を重ね、認識率の評価と認識結果の処理アルゴリズムの改良を行っていく。

参考文献

- (1) 松井俊弘ほか：文字認識技術コンテスト実施結果，手書き文字認識技術の過去・現在・未来シンポジウム講演論文集，電子情報通信学会，pp.15(1993)
- (2) 福戸淳司ほか：音声入出力を用いた一名当直時の人間を含むシステムの異常対策について，日本航海学会論文集，第102号，pp.63~70(2000)
- (3) 大山敏夫ほか：音声指令・ガイドによる航海支援システム，三菱重工技報，第197号，pp.400~403(1996)
- (4) 全国内航タンカー海運組合ほか：航海支援システム開発に関する共同研究報告書(1998)
- (5) 全国内航タンカー海運組合ほか：OMBO／航海支援システム成果発表会議事録(1998)

質疑応答

村山義夫（海上労働科学研究所）：

人が発する言葉は辞書の用語と完全に一致した場合に有効なのか、わずかな違いでも許容されるのか教えて下さい。

松田和生：

辞書に登録されている用語を発声しないと認識されません。わずかな違いがある場合にも、基本的には認識できません。

嶋田博行（神戸大学）：

認識率とエラー率を足すと100%になっていませんが、この関係について教えて下さい。また、緊急の場合、誤動作の対応についてどのように考えられていますか。

松田和生：

認識結果としては、正しい認識、棄却、エラーの3種類で評価しています。棄却は、一つの認識結果として決定できない場合で、(正しい)認識率、棄却率、エラー率を合計すると100%となります。また、緊急時の誤動作についてですが、画面に認識結果や次のアクションに対する操作ウィンドウを表示することにより、対処を急ぐ場合には手でもオペレーションできるように考慮しています。