電力系統監視制御のための音声入力インタフェース

非会員 瀬 川 修 (中部電力株式会社)
非会員 高 田 亨 (中部電力株式会社)
非会員 高 橋 誠 (株式会社東洋情報システム)
非会員 西 俣 憲 幸 (株式会社東洋情報システム)
非会員 有 本 雅 昭 (三菱電機株式会社)

Development of Speech-Input Interface for Power Control and Information System

Osamu Segawa, Non-member, Tooru Takata, Non-member (Chubu Electric Power Co.,Inc.), Makoto Takahashi, Non-member, Noriyuki Nishimata, Non-member (Toyo Information Systems Co.,Ltd.), Masaaki Arimoto, Non-member (Mitsubishi Electric Corporation)

For utilization in power control and information systems, we have developed human-computer interfaces using speech recognition technologies. For the purpose of more effective console operation, we propose a method of speech-input interface using voice navigation commands. In a large scale power control center, the system console has a large number of menu screens and the layer of console menu is very complicated. Therefore, the proposed method is very effective for console operation because the interface enables rapid, hands-free, eyes-free operation. For implementation of the speech-input interface, we have developed a large vocabulary speech recognition module and a system integration method between the speech-input interface and power control center's host computer systems. In this paper, we describe initial implementation and evaluation of the proposed speech-input interface.

キーワード：音声認識, ヒューマン・インタフェース, 電力系統監視制御システム

1. はじめに

電力系統の監視制御の分野では、システムのインタフェース機能の重要性は今後も高まる一方であり、系統状態の迅速かつ的確な把握や、運転操作時の人為的ミスに対する安全性の向上が期待される。これらの課題に対し、従来は合成メッセージ出力による警報など一部の限定された用途に採り入れられてきた音声メディアをインタフェースとして積極的に利用することが考えられる⁽⁴⁾⁽⁵⁾。音声認識技術を応用したインタフェースは、画面のメニュー選択において視覚や直接操作によらない迅速な画面・機能呼び出しが可能であるなど、ヒューマン・インタフェースの実現方式として有効であると言われている⁽⁶⁾⁽⁷⁾。

電力分野においても、今後の大規模システムのオペレーションに求められる機能として(1)操作ステップ数の短縮、(2)視覚への負担軽減などがあり、音声入力インタフェースの導入はこれらの課題に対し有効な方策であると考えられる。これまで筆者らは音声入力インタフェースによる画面オペレーションの実現を目的として、音声認識モジュール

の実装⁽¹⁾と試作システムによる機能検証を中心に検討を行ってきた。しかしながら、新しい概念の入力デバイスを実システムへ組み込むためには、制御用ホスト計算機や操作卓との連携方式をはじめ、オペレータが使用する入力フロントエンドのインタフェース実装など、従来の簡易なデモシステムの枠組から脱した詳細な実現方式の検討が必要である。これまで、電力をはじめとする大規模監視制御システムのオペレーションを目的として、音声入力インタフェースを実システムに組み込んで機能検証と評価を行なった事例は少ない。そこで、本研究では音声入力インタフェースの次世代システムでの実現を目標として、音声認識モジュールの実装と制御所 LAN による制御ホスト計算機との連携方式などの検討を行なった。

以下、本論文では提案する音声入力インタフェースの機能を説明し、これらを実現する要素技術として音声認識処理の手法を述べる。さらに、音声入力インタフェースを既設システムに組み込み、制御用ホスト計算機と連携させる方式を述べる。そして、提案方式を工場試験中の訓練システムに実装を行ない、連携方式の評価検証を行なった結果

を報告する。

2. 音声入力インタフェースの概要

提案するインタフェースは、給電制御所の操作卓(以下、制御卓 EWS と記す)での使用を目的としている。対象となるユーザは電力会社の社員であり業務に関して専門知識を持ったオペレータである。給電制御所の通常業務の主なものは、(1)電力系統の状態監視と(2)電力設備の保守運用に伴う現地設備の遠隔操作である。本研究では、音声入力インタフェースの適用箇所として以下の「音声コマンドによる画面オペレーション」を設定した。

〈2・1〉音声コマンドによる画面オペレーション 制御卓 EWS には入力装置としてポインティングデバイス(スタイラスペンまたはマウス)とボタンが備わっている。オペレータは通常の画面オペレーションではポインティングデバイスまたはボタンを用いている。音声コマンドによる画面オペレーションはこれら従来の入力デバイスの機能の一部を、音声入力インタフェースによって置き換えるものである。

図1、図2に実システムの送電線リストと画面体系の一例を示す。制御所の規模が大きくなるほど操作画面やメニューの選択項目は多くなり、送電線単結など各種画面や機能呼び出しを迅速かつ正確に行う機能が必要とされる。選択項目が視覚で一度に捉え切れない、画面内に表示されていない、あるいは一つの画面に収まらない程度に選択肢が多い場合(例えば、図2の第2階層リスト画面は「あ、か行」など送電線の読みによって3画面に分割されている)は音声による選択のメリットがある。さらに音声と他の直接操作による入力手段は併用が可能であることから、その同時性・共働性を活かした効率向上が期待できる。

〈2・2〉画面オペレーションの機能 音声入力インタフェースによって次の2つの機能を実現する。

音声アイコン(多項目検索): 画面上で対象の直接操作による選択が困難な場合(両手が塞がっている、選択対象が隠れている、視覚的に探しにくいなど)音声コマンドによって間接的に選択を行なう機能である。本機能では、ボタンやポインティングデバイスによって直接選択する代わりに、目的の画面(リスト画面、送電線・電気所の単結画面、および各種機能の画面)の名称を音声コマンドとして発声することにより画面呼び出しを行なう。

音声ショートカット(操作ステップ短縮): 画面展開が深い階層まである時に、目的の画面を音声コマンドによって一つのアクションで呼び出す機能である。

3. 音声認識システムの実現方式

〈3・1〉認識タスクの設定 タスクの規模と特徴を把握するために、実システムの画面体系と一画面の選択項目の数を調査した。その結果、画面ごとの選択項目(アイコン)の数は、「送電線」や「電気所」のリスト画面など多い箇所

で30~50程度、主要な画面体系の階層の深さは2~3層程

送電線単結リスト 33kV (あ、か行)		
1 ■ 愛知時計プラザー港線	2 3 ■ 水主町小山線	4 5 ■ ...
2 ■ 稲永愛知機械永徳線	2 4 ■ 水主町八郎線	4 6 ■ ...
3 ■ 稲永稲永埠頭線	2 5 ■ 鳥森下広井線	4 7 ■ ...
4 ■ 稲永新日本製鉄線	2 6 ■ 鳥森交通局岩塚線	4 8 ■ ...
5 ■ 稲永東邦ガス空見線	2 7 ■ 鳥森交通局本陣線	4 9 ■ ...
6 ■ 稲永宝神下水線	2 8 ■ 鳥森二ツ橋線	...
:	:	:

図1 送電線リスト画面の一例
Fig. 1. An example of powerline list menu.

図2 画面体系の一例
Fig. 2. An example of screen menu structure.

度であり、一度の認識カテゴリの数は300~500語程度必要であることがわかった。また発声単位であるが、画面オペレーションは自発的発声であることから、システム側で制約を与えない限り言語的拘束による知識を認識システムに組み込みにくいと言える(逆にユーザ側も、どのように発声したら意図が達成されるのかが明確でない)。本システムではオペレータの発声方法に関する制約を設定し、あらかじめコマンドとして定義した離散単語または連続文を受理可能とした。

〈3・2〉音声認識モジュールの実装 文法の記述によって連続音声の認識も可能なオートマTON制御One-Pass-DP法によるシステムを作成した。システムの諸元については〈5・1〉で後述する。本システムは汎用UNIXワークステーション上で動作し、特殊なハードウェア類を必要としないソフトウェア認識システムである。本研究の対象は電力会社内部での使用を前提としたシステムであるため、ユーザは小人数の話者に限定されており、また音声入力インタフェースに対する訓練・習熟期間を設けることが十分に可能である。このことから本システムでは不特定話者認識をベースとして、移動ベクトル場平滑化方式(VFS法)⁽³⁾による話者適応機能を併用している。

〈3・3〉音素トライ構造辞書 本タスクでは一度の認識カテゴリの数が多いため、単純な辞書構造では音響モデル(音素HMM)の照合時の処理効率がよくない。また認識カテゴリである単語(送電線名、電気所名、各種機能名)の読みの

特徴として、「潮流図」^{ちようりゅうず}、「貯水池運用計画」^{ちようすいちうんようけいかく}、「中濃犬山」^{ちようのういぬやま}、「中濃蘇原」^{ちようのうそはら}、「中濃関」^{ちようのうせき}のように接頭部を音素単位で共有できる箇所が少なからず存在する。そこで、本システムでは図3に示すような音素トライ構造辞書を実装した。図3の例は上述の5つの単語を音素トライ構造で記述したものであり、太線の音素ノードの照合を共通に進めることができる。

図3 音素トライ構造辞書

Fig. 3. An example of phoneme TRIE structure dictionary.

4. 制御ホスト連携の実現方式

ここでは、制御所システムに追加モジュールとして音声認識サーバを組み込み、既設の制御卓EWSのソフトウェアのみの改修によって音声入力インタフェースの実装を可能にする方式を述べる。

本研究で連携の対象とする制御所システムは現行で最も新しい世代のシステムであり、制御卓は汎用UNIXワークステーションをベースに電力向にハードウェアとOSをカスタマイズしたものである。システムの各ノードはイーサネットLANに接続されている。対象システムは、基本的に標準のUNIX機のアーキテクチャおよび通信プロトコルを実装していることから、モジュールの追加と通信による連携は比較的容易であるとの見通しがあった。このため、既設モジュールの中で制御卓EWSの最小限の改造によって音声入力インタフェースの機能を付加できる方式を考案した。

〈4・1〉連携方式の概要 連携方式の概要を図4に示す。モジュール追加・改修による主要な変更点を以下に示す。

音声認識サーバAPI: 音声認識サーバと制御卓EWSの連携のために通信APIを作成した。サーバで使用する音声認識システムはタスクに依存しない汎用的なモジュール構成になっているが、制御所システムで用いる場合は通信APIを組み込むだけで利用可能となる。音声認識サーバは送信関数をコールすることによって認識結果コード(画面ID)を制御卓EWSのプログラムに送信する。

イベント管理: 既設入力デバイスのイベント(ポインティング、ボタンの押下)と音声入力インタフェースのイベントの両方を受け付けるために、イベント管理プログラムの割り込み制御部分の改修を行なった。

ショートカットによる画面遷移: 制御卓EWSでは受信した認識結果コードを変換テーブルを用いてオペレーシ

ョン管理プログラムが処理できる画面制御コマンドに変換する。オペレーション管理では既設入力デバイスと音声入力のイベントは同等に扱われるが、画面制御に関しては本来1階層ごとの遷移しか考慮されていない(すなわちショートカットによる画面遷移については対応するイベント定義がない)。そこで、音声ショートカットを実現するために目的画面ごとにあらかじめ定義した表示ルートイベント列(例えば、親画面→子画面→目的画面)をテーブルに保持しておき、これらのイベント列を一度に実行することによって途中の画面表示を飛ばした遷移を実現した。

図4 連携方式の概要

Fig. 4. Overview of the system integration method.

〈4・2〉音響入力フロントエンドの実装 音響入力フロントエンドは、ユーザが制御卓EWSから音声コマンドを入力するためのインタフェース・プログラムである。本研究では制御卓EWSの負荷を軽減し実装が容易な方法を検討した結果、音声認識サーバ側で生成した音響入力フロントエンドのウィンドウを、X-WindowのXプロトコルによって制御卓EWSの画面にリモート表示する方式を用いている。音響入力フロントエンドのウィンドウは制御卓画面の空きスペースに表示され、音声入力レベルのインジケータや音声認識サーバの処理状態などのシステム情報も併せて表示される(図5参照)。

図5 音響入力フロントエンドのウィンドウ

Fig. 5. Audio input front-end interface window.

実際の音声入力的方式であるが、図6に示すとおりマイク ON だけはユーザがマウスまたはスタイラスペンのクリックによって明示的にシステムに通知し、入力音声の始端・終端の検出と切り出しは自動で処理している。また認識システムの内部状態をユーザにわかりやすく開示する表示ランプを設けることによって、インタフェースとしての使用感を向上させている。

図6 音響入力フロントエンドの処理

Fig. 6. Process flow of the audio input front-end interface.

〈4.3〉 特殊なナビゲーション機能

〈4.3.1〉 表示先コンソール画面の指定 対象システムに固有の仕様として、各制御卓は3つのコンソール画面を備えている。このため音声入力による画面ナビゲーションにおいても表示先コンソールを指定する機能は利便性が高いと考えられる。そこで表示先コンソール画面(左, 中, 右)を連続文の音声コマンドで指定できるナビゲーション機能を実現した。連続音声認識の文法は図7に示すような有限状態オートマトン(正規文法)で記述する。

図7 連続音声認識の文法の記述

Fig. 7. An example of grammar for continuous speech recognition.

〈4.3.2〉 音声コマンドによる追加機能の実現 音声コマンドとシステムの内部コマンド定義のバインディングによって既設デバイスに無いナビゲーション機能を容易に追加することが可能である。例えば、通常のシステムでは一つ前の表示画面に遷移するナビゲーション機能はハードウェアやGUIのボタンとして実装されていないが、「戻る」という音声コマンドと画面制御イベントの組合せによって、

既設のコンソールを物理的に改造することなく追加機能を実現することができる。

5. 音声認識モジュールの基本性能評価

ここでは実装した音声認識モジュールの概要と基本性能評価について述べる。

〈5.1〉 音声認識モジュールの諸元 音響分析条件は、サンプリング周波数 12kHz, 16bit 量子化, 20ms ハミング窓, フレーム周期 10ms, 特徴パラメータとして, 16 次 LPC ケプストラム, 16 次 Δ LPC ケプストラム, 対数パワー, Δ 対数パワーの合計 34 次元を用いた。音響モデルは音素環境依存モデル HMnet⁽²⁾を用いた。

〈5.1.1〉 実験条件 ATR 音声データベースを用いて生成した不特定話者モデル(ベースモデル, HMnet 400 状態, 5 話者混合)に対し, 男性話者 1 名が発声したタスクの音声データ 235 単語を用いて, VFS 法⁽³⁾により話者適応を行なった適応モデルを作成した。

〈5.1.2〉 単語認識率の評価 実験は画面オペレーションタスクの送電線名, 電気所名など 235 単語を辞書登録した単語認識で評価を行なった。評価データとして男性話者 1 名のタスク 235 単語の音声データ(音響モデルの学習に用いていないオープンデータ)を用意し, ビーム幅を変えて評価を行なった。単語認識率を表1に示す。本モジュールは画面オペレーションタスクで使用するのに十分な性能を有している。

表1 単語認識による基本性能評価

Table 1. Word recognition rate of the baseline system.

ビーム幅	300	500	800
ベースモデル	69.3(%)	83.8(%)	90.6(%)
適応モデル	92.8(%)	97.9(%)	99.6(%)

6. システム連携方式およびインタフェースの評価

〈6.1〉 検証内容 工場内で試験中の訓練シミュレータのセット(制御ホスト計算機1台と制御卓 EWS4 台より構成, 実システムと同様の機能や画面データを持つ)に音声入力インタフェースを実装し, 提案方式によるシステム連携とインタフェースの有用性を検証した。

音声入力で使用可能なコマンドは, 送電線, 電気所のリスト画面, 単結画面の名称, およびコンソール・ボタンで表示できる各種機能画面の呼び出しなど合計 235 種類と, 〈4.3〉で述べた特殊ナビゲーション・コマンドを設定した。実験では音声認識サーバとして SPECint95 6.26 の処理性能の WS を使用し, 音声入力のマイクロフォンは両手が自由に使えるヘッドセットを用いた。

〈6.2〉 基本機能の検証 制御卓 4 台の内の 1 卓を使用して音声入力インタフェースによるオンラインの画面オペレーションを行ない, 提案方式によるシステム連携が良好に動作することを確認した。音響入力フロントエンドに

についても入力の実安定性や発話タイミングの取り易さを確認することができた。

〈6・3〉 インタフェースの有用性評価 音声入力インタフェースの有用性を評価するために、画面オペレーションタスクを設定して従来デバイス（ボタン、スタイラスペン）との比較実験を行なった。インタフェースとしての効率性を示す尺度としては、次の2点により評価する。

- 操作所要時間
- 操作ステップ数

評価タスクは、送電線単結、電気所単結など、図2の深い階層の画面（孫画面）の呼び出しである。ここで測定する操作所要時間は、音声入力インタフェースの場合はコマンドの発声が完了してから画面表示が完了するまでの時間である（すなわち、音声認識処理とホスト側処理の合計時間）。従来デバイスの場合は、試験監督者の発声による画面名指示が完了してから画面表示が完了するまでの時間である（すなわち、ボタン・スタイラスペン操作時間とホスト側処理の合計時間）。

〈6・3・1〉 効率性の評価 音声入力インタフェースについては専門オペレータではない被験者3名で、単語コマンドによる画面オペレーションの試行100回によって評価した。従来デバイスについては被験者1名（対象システムの熟練オペレータ）で、同一タスクの試行100回によって評価した。音声入力インタフェースの平均操作所要時間を表2に示す。音声入力インタフェースの場合の操作ステップ数はいずれも1回である。

表2 音声入力インタフェースの効率性評価
Table 2. Result of average operation time using the speech-input interface.

	被験者1	被験者2	被験者3
平均操作所要時間(秒)	9.55	9.76	8.75

従来デバイスの平均操作所要時間と平均操作ステップ数を表3に示す。

表3 従来デバイスの効率性評価
Table 3. Result of average operation time and step number using the former interface.

平均操作所要時間(秒)	10.74 (最大31秒)
平均操作ステップ数(回)	2.51 (最大6回)

〈6・4〉 実環境での音声認識性能評価 検証場所の実環境下でオンラインによるタスク100単語の認識率を測定した。辞書登録235単語、音響モデルは〈5・1・2〉のベースモデルに対して各被験者ごとに話者適応（被験者1は235単語、被験者2は16単語、被験者3は16単語）を行なったものを用い、ビーム幅は800に設定した。実環境での単語認識率を表4に示す。

被験者2と被験者3の認識率がかなり低下しているのは、話者適応のサンプル単語が16語と十分でなかったことにも

表4 実環境での単語認識率

Table 4. Result of word recognition rate at the real environment.

被験者1	被験者2	被験者3
97.0(%)	85.0(%)	90.0(%)

起因するが、主要因は検証場所の工場内の音環境が騒音レベル61～66dB(A)とかなり悪かったことである。周辺の計算機類のノイズ（定常雑音）によって、音声区間の検出誤りなどが生じたものと考えられる。単発的な周囲の人の声やアラーム音については、指向性のあるヘッドセットを用いることではば対処できることが確認できた。なお実際の制御所のオペレーション室は騒音レベル約50dB(A)で実験室レベルに近い比較的静かな音環境である。

〈6・5〉 考察と音声認識モジュールの高速化 まず操作ステップ数については、音声入力インタフェースは深い階層の画面でも確実に1ステップで呼び出すことができるため、従来デバイスよりも効率が良い。操作所要時間については、従来デバイスのオペレーションでは遅い時（普段あまり使わない画面などの表示）は最大30秒以上を要するため、音声入力インタフェースの方が時間的に安定しているという結果であった。ただし操作所要時間の平均値に関しては従来デバイスと同等程度で音声入力インタフェースの明確な優位性を示すまでには至らなかった。これは音声認識モジュールの処理が十分に高速でないことが原因である。そこで、認識処理のスコア計算など実装部分の見直しによって高速化した新しい認識モジュールと、さらに高速なWS (SPECint95 13.0)を用いて操作所要時間の評価を再度行なった。被験者は1名（表2の被験者1と同じ）で、画面オペレーションの試行回数は、単語コマンド、連続文コマンドともに各20回である。実験結果を表5に示す。

表5 高速化認識モジュールを用いた再評価
Table 5. Result of average operation time using high performance speech recognition module.

	単語コマンド	連続文コマンド
平均操作所要時間(秒)	4.66	5.55

高速化した音声認識モジュールを用いた場合、平均操作所要時間は従来デバイスの10.74秒に対し単語コマンドで4.66秒と音声入力インタフェースの優位性を十分に示す結果が得られた。表示先コンソール指定など連続文コマンドについても単語コマンドとは遜色ない処理時間を実現している。

7. おわりに

本論文では音声認識技術の電力分野への応用として、音声メディアを積極的に利用する新しい概念のインタフェースの提案を行なった。音声入力インタフェースが有効に使用できる業務タスクとして「音声コマンドによる画面オペレーション」を設定し、これを実現する音声認識処理の手

法と実システム組み込みのための連携方式について述べた。

提案方式を訓練シミュレータに実装し、フィールド試験を行なった結果は次のとおりである。

- 音声入力インタフェースによる操作性向上を確認した。音声入力は視覚に負担が少なく、操作所要時間、操作ステップ数の短縮に有効である。
- 音声認識機能のモジュール化、制御所 LAN による簡易な連携方式など、システムの最小限の改修によって音声入力インタフェースを実装する方式を提案し、検証試験によって連携機能の実現性を確認した。

なお現在実施している実証研究の範囲では、まだ実運用システムへの導入予定はないが、将来的には次世代システムの標準インタフェースとしてシステム仕様に盛り込むことを目標にしている。また音声認識モジュールについてもアルゴリズムの改良によって、さらに認識性能、ロバスト性、リアルタイム性の向上を図る予定である。

(平成 10 年 10 月 23 日受付, 同 11 年 5 月 6 日再受付)

文 献

- (1) 瀬川修, 高橋誠, 水本綾: 「電力系統監視制御のための音声入力インタフェースの検討」, 人工知能学会第 11 回全国大会論文集, pp.396-399 (1997)
- (2) 鷹見淳一, 嵯峨山茂樹: 「逐次状態分割法による隠れマルコフ網の自動生成」, 信学論 (D-II), Vol. J76-D-II, No.10 pp.2155-2164 (1993)
- (3) 大倉計美, 杉山雅英, 嵯峨山茂樹: 「混合連続分布 HMM を用いた移動ベクトル場平滑化話者適応方式」, 信学技報, SP92-16 (1992)
- (4) 藤井光彦, 石川泰, 寺下尚孝, 須山勉, 佐伯昭, 栗田和彦: 「原子力発電プラント マンマシンシステム (MMS)II - マンマシンシステムの要素技術 -」, 三菱電機技報, Vol.63, No.7 (1989)
- (5) 三宅雅夫, 谷越浩一郎, 畑岡信夫, 大賀幸治, 宗政成大, 小金宏行: 「人にやさしい監視制御システムのヒューマンインタフェース」, 日立評論, Vol.78, No.10 (1996)
- (6) 日本情報処理開発協会編: 「音声の知的処理に関する調査研究」, (財) 日本情報処理開発協会 (1993)
- (7) C.Schmandt: 「コンピュータとのヴォイスコミュニケーション」, 石川泰 訳, サイエンス社 (1995)

瀬川 修 (非会員) 1991 年横浜国立大学工学部電子情報工学科卒業。1993 年同大学院工学研究科博士課程前期修了。同年中部電力株式会社入社。以来、電力技術研究所において主に音声認識、自然言語処理に関する研究に従事。情報処理学会, 日本音響学会各会員。

高田 亨 (非会員) 1985 年大阪大学工学部電気工学科卒業。同年中部電力株式会社入社。系統運用業務を経て、現在同社電力技術研究所にて電力系統運用計画に関する研究に従事。

高橋 誠 (非会員) 1989 年大阪外国語大学外国語学部中国語学科卒業。同年株式会社東洋情報システム入社。以来、言語処理グループに所属し、自然言語処理、機械翻訳システム、音声認識システムの研究開発に従事。人工知能学会会員。

西 俣 憲 幸 (非会員) 1994 年大阪教育大学教育学部卒業。1996 年大阪教育大学大学院教育学研究科修了。同年株式会社東洋情報システム入社。以来、自然言語処理、音声認識システムの研究開発に従事。電気情報通信学会会員。

有本 雅 昭 (非会員) 1990 年 甲南大学理学部応用数学科卒業。同年 三菱電機株式会社入社。以来、電力系統部において、主に系統制御システムの基本 S/W (ミドルウェア) に関する開発に従事。